Contemporary Mathematics

Application of Logic to Combinatorial Sequences and Their Recurrence Relations

Eldar Fischer, Tomer Kotek, and Johann A. Makowsky

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, or equivalently, the generating function $g_L(x) = \sum_n d_L(n)x^n$ is a rational function.

In this paper we survey results concerning sequences $a(n)$ of natural numbers which (i) have a combinatorial or logical interpretation and (ii) satisfy linear recurrence relations over $\mathbb{Z}$ or $\mathbb{Z}_m$. We discuss the pioneering, but little known, work by C. Blatter and E. Specker from 1981, and its further developments, including results by I. Gessel (1984), E. Fischer (2003), and related results by the authors.

Version June 28, 2011

2010 *Mathematics Subject Classification.* 03-02, 03C98, 05-02, 05A15, 11B50 .

Partially supported by an ISF grant number 1101/06 and an ERC-2007-StG grant number 202405.

Partially supported by the Fein Foundation and the Graduate School of the Technion - Israel Institute of Technology.

Partially supported by the Israel Science Foundation for the project "Model Theoretic Interpretations of Counting Functions" (2007-2010) and the Grant for Promotion of Research by the Technion–Israel Institute of Technology.

©2009 American Mathematical Society

Contents

Part 1. Introduction and Synopsis

1. Sequences of integers and their combinatorial interpretations

We discuss sequences $a(n)$ of natural numbers or integers which arise in combinatorics. In some cases such sequences satisfy linear recurrence relations with constant or polynomial coefficients. In this paper we discuss sufficient structural conditions on $a(n)$ which imply the existence of various linear recurrence relations.

The traditional approach for studying such sequences consists of interpreting $a(n)$ as the coefficients of a generating function $g(x) = \sum_n a(n)x^n$, and of using *analytic methods*, to derive properties of $a(n)$, cf. [**34**]. Another general framework of *species of structures* for combinatorial interpretations of counting functions is described in [**8, 9**]. Lack of space does not allow us to use this formalism in this paper. There is also a substantial theory of how to algorithmically verify and prove identities among the terms of $a(n)$, see [**57**].

We are interested in the case where the sequence $a(n)$ admits a *combinatorial* or a *logical* interpretation, i.e., $a(n)$ counts the number of some relations or functions over the set $[n] = \{1, \dots, n\}$ which have a certain property possibly definable in some logical formalism (with or without its natural order). To make this precise, we assume the reader is familiar with the very basics of Logic and Finite Model Theory, cf. [**25, 49**]. We shall mostly deal with the logics **SOL**, Second Order Logic, and **MSOL**, Monadic Second Order Logic. Occasionally, we formulate statements in the language of automata theory and regular languages and use freely the Büchi-Elgot-Trakhtenbrot Theorem, which states that a language is regular iff it is definable in **MSOL** when we view its words of length n as ordered structures over a set of n elements equipped with unary predicates, cf. [**25**]. More details on the logical tools used will be given wherever needed.

We define a general notion of *combinatorial interpretations* for finite ordered relational structures.

Definition 1.1 (Combinatorial interpretation). *A* combinatorial interpretation $\mathcal{K}$ of $a(n)$ *is given by*

(i) a class of finite structures $\mathcal{K}$ over a vocabulary $\tau = \{R_1, \dots R_r\} = \{\bar{R}\}$ or $\tau_{ord} = \{<_{nat}, \bar{R}\}$ with finite universe $[n] = \{1, \dots, n\}$ and a relation symbol $<_{nat}$ for the natural order on $[n]$.

(ii) The counting function $d_{\mathcal{K}}(n)$, which counts the number of relations[1]

$$d_{\mathcal{K}}(n) = \mid \{\bar{R} \text{ on } [n] : \langle [n], <_{nat}, \bar{R}\rangle \in \mathcal{K}\} \mid$$

such that $d_{\mathcal{K}}(n) = a(n)$.

(iii) A combinatorial interpretation $\mathcal{K}$ is a pure combinatorial interpretation *of $a(n)$ if $\mathcal{K}$ is closed under τ-isomorphisms. In particular, if $\mathcal{K}$ does not depend on the natural order $<_{nat}$ on $[n]$, but only on τ.*

The way we defined combinatorial interpretations does not require *uniformity*. In the ordered case $\mathcal{K}$ could be patched together arbitrarily, in the pure case we only have to require that it is closed under isomorphisms. Uniformity can be formulated in terms of some device (a Turing machine, a logical formula) or closure conditions

[1]In enumerative combinatorics there are various terminologies. If the counting function is monotone, it is called speed in [**5**].

(closed under substructures, products, disjoint unions). In this article we are mostly concerned with the case of logical formulas. Intuitively speaking, a combinatorial interpretation $\mathcal{K}$ of $a(n)$ is a *logical interpretation of* $a(n)$ if $\mathcal{K}$ is definable by a formula in some logic formalism, say full Second Order Logic.

DEFINITION 1.2 (Logical interpretation and Specker sequences).

(i) *A combinatorial interpretation* $\mathcal{K}$ *of* $a(n)$ *is an* **SOL**-interpretation (**MSOL**-interpretation) of $a(n)$, *if* $\mathcal{K}$ *is definable in* $\mathbf{SOL}(\tau_{ord})$ *(*$\mathbf{MSOL}(\tau_{ord})$*).*
(ii) Pure **SOL**-interpretations (**MSOL**-interpretation) of $a(n)$ *are defined analogously.*
(iii) *We call a sequence* $a(n)$ *which has a logical interpretation in some fragment* $\mathcal{L}$ *of* **SOL** *an* $\mathcal{L}$-Specker sequence*, or just a* Specker sequence *if the fragment is* $\mathbf{SOL}$[2].

The following two propositions are straightforward.

PROPOSITION 1.1.

(i) *If* $a(n)$ *has a combinatorial interpretation then for all* $n \in \mathbb{N}$ *we have* $a(n) \geq 0$.
(ii) *If* $a(n)$ *has a combinatorial interpretation then for all* $n \in \mathbb{N}$ *we have* $a(n) \leq 2^{n^{d(\tau)}}$, *where* $d(\tau)$ *is a constant depending on the vocabulary* τ.
(iii) *There are uncountably many sequences* $a(n)$ *which have a combinatorial interpretation.*

PROOF. We only prove (iii). Let $\mathcal{K}_0$ be the class of structures with one unary predicate of the form $\langle [n], \emptyset \rangle$. Let $\mathcal{K}_1$ be the class of structures with one unary predicate of the form $\langle [n], P \rangle$. We have $d_{\mathcal{K}_0}(n) = 1$ and $d_{\mathcal{K}_1}(n) = 2^n$. Now let $A \subseteq \mathbb{N}$ and define

$$\mathcal{K}_A = \{\langle [n], \emptyset \rangle : n \in A\} \cup \{\langle [n], P \rangle : n \notin A\}$$

Clearly, for each $A \subseteq \mathbb{N}$, the class $\mathcal{K}_A$ gives a pure combinatorial interpretation for the sequence

$$d_{\mathcal{K}_A}(n) = \begin{cases} 1 & n \in A \\ 2^n & n \notin A \end{cases}$$

□

PROPOSITION 1.2.

(i) *There are only countably many Specker sequences.*
(ii) *Every Specker sequence* $a(n)$ *is computable, and in fact it is in the counting class* $\sharp \cdot \mathbf{PH}$, [**42**], *where* **PH** *is the polynomial hierarchy and the input* n *is measured in unary presentation. Hence* $a(n)$ *is computable in exponential time, and using polynomial space.*
(iii) *The set of Specker sequences is closed under the point-wise operations of addition and multiplication. The same holds for* **MSOL***-Specker sequences.*

[2]E. Specker was to the best of our knowledge the first to introduce **MSOL**-definability as a tool in analyzing combinatorial interpretations of sequences of non-negative integers.

We shall discuss examples of sequences $a(n)$ which have a combinatorial interpretation in great detail in Part 2. Among them we find the counting functions for trees, graphs, planar graphs, binomial coefficients, factorials, and many more. The reader may want to consult the *The On-Line Encyclopedia of Integer Sequences (OEIS)* [**1**], which contains close to 200'000 integer sequences which were studied in the literature. The non-monotonic sequence $a(n)$ beginning with

$$\begin{gathered} 1,1,1,2,5,4,7,6,7,9,11,10,13,13,13,14, \\ 17,16,19,18,19,21,23,22,25,25,25,26,29, \\ 28,31,30,31,33,35,34,37,37,37,38,41, \\ 23,22,25,25,25,26,29,28,31,30,31,33, \\ 35,34,37,37,37,38,41,40,43,42,43,45,47, \\ 46,49,49,49,50,53,52,55,54,55,57,59,58,\ldots \end{gathered} \tag{A085801}$$

appears there as Sequence A085801 which has an ordered combinatorial interpretation as the *maximum number of nonattacking queens on an* $(n \times n)$ *toroidal board.*

To the best of our knowledge no systematic study of sequences $a(n)$ which have a combinatorial interpretation has been undertaken so far. A notable exception are the cases were $\mathcal{K}$ is a graph property which is hereditary (closed under induced subgraphs) or monotone (closed under subgraphs), cf. [**64, 5, 6, 4**].

Problem 1.

(i) Characterize the sequences of integers which have (pure) combinatorial interpretations under various restriction imposed on $\mathcal{K}$. What are the additional restrictions on the counting function besides those listed in Proposition 1.1?

(ii) Characterize the $\mathcal{L}$-Specker sequences for various sublogics of **SOL**.

However, in this paper we investigate sufficient conditions for Specker sequences to satisfy linear recurrence relations. We also study the converse question: For a given type of linear recurrence relations REC, can we find a family of logical interpretations $\mathcal{I}$ such that every sequence satisfying REC has a logical interpretation in $\mathcal{I}$?

2. Linear recurrences

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

Definition 2.1 (Recurrence relations). *Given a sequence $a(n)$ of integers we say $a(n)$ is*

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

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

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

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

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

where each p_i *is a polynomial in* $\mathbb{Z}[X]$ *and* $p_q(n) \neq 0$ *for any* n. *We call it* Simply-P-recursive or SP-recursive, *if additionally* $p_q(n) = 1$ *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) MC-finite *(modularly C-finite), if for every* $m \in \mathbb{N} \backslash \{0\}, m > 0$ *there is* $q(m) \in \mathbb{N} \backslash \{0\}$ *for which* $a(n)$ *satisfies for all* $n > q(m)$

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

where $q(m)$ *and* $p_i(m)$ *depend only on* m, *and* $p_i(m) \in \mathbb{Z}$. *Equivalently,* $a(n)$ *is MC-finite, if for all* $m \in \mathbb{N}$ *the sequence* $a(n) \pmod m$ *is ultimately periodic.*

(v) trivially MC-finite, *if for each* $m \in \mathbb{N}$ *and large enough* n, $f(n) \equiv 0 \pmod m$.

The terminology C-finite and holonomic are due to [**75**]. P-recursive is due to [**68**]. P-recursive sequences were already studied in [**11, 12**].

The following are well known, see [**34, 28**].

Lemma 2.1.

(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, MC-finite, SP-recursive and P-recursive 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 any integer m, [**36**].

Proposition 2.2. *Let* $a(n)$ *be a function* $a : \mathbb{N} \to \mathbb{Z}$.

(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 SP-recursive then* $a(n)$ *is MC-finite.*

(iv) If $a(n)$ *is hypergeometric then* $a(n)$ *is P-recursive.*

Moreover, the converses of (i), (ii), (iii) and (iv) do not hold, and no implication holds between MC-finite and P-recursive.

Proof. The implications follow from the definitions.
$n!$ is SP-recursive, but not C-finite, as, by Lemma 2.1 it grows too fast.
$\frac{1}{2}\binom{2n}{n}$ is P-recursive but not MC-finite, see the discussion in Example 7.4.
The Bell numbers $B(n)$ are MC-finite, but not P-recursive, hence not SP-recursive,

see the examples in Example 7.5.
The derangement numbers $D(n)$ in Example 7.3 are SP-recursive but not hypergeometric, cf. [**57**]. To see that no implication holds between MC-finite and P-recursive, note that n^n is MC-finite, but not P-recursive. Furthermore, $\frac{1}{2}\binom{2n}{n}$ is P-recursive, but not MC-finite, see the Examples 7.4 and 7.2. □

REMARKS 2.1. *There are sequences with pure combinatorial interpretations which are neither MC-finite nor P-recursive. As an example, take the sequence* $\frac{1}{2}\binom{2n}{n} + 2n^{2n}$. *It counts the number of binary relations* $R \subseteq [2n]^2$ *such that* R *is either the edge relation of a graph which consists of two cliques of equal size, or* R *is a function* $R : [2n] \to [2n]$.

PROPOSITION 2.3.

(i) *There are only countably many P-recursive sequences* $a(n)$.
(ii) *There are continuum many MC-finite sequences.*
Furthermore, let $a(n)$ *be an MC-finite sequence such that* $a(n) = 0 \pmod m$ *for all* m *and* $n \geq q(m)$, *and* $a(n)$ *is nonzero for all large enough* n. *For* $A \subseteq \mathbb{N}$, *let*

$$a_A(n) = \begin{cases} a(n) & n \in A \\ 2 \cdot a(n) & n \notin A \end{cases}$$

Then $a_A(n) = 0 \pmod m$ *for all* m *and* $n \geq q(m)$.

3. Logical formalisms

3.1. Fragments of SOL. Let $\bar{R}$ be a finite set of relation symbols. First Order Logic (**FOL**) over $\bar{R}$ has the atomic formulas "$R_i(x_1, \ldots, x_{\rho(i)})$" and "$x_1 = x_2$" where $x_1, x_2, \ldots$ are any individual variables. The set of formulas $\mathbf{FOL}(\bar{R})$ denotes all formulas composed from the atomic ones using boolean connectives, and quantifications of individual variables "$\exists x$" and "$\forall x$".

Second Order Logic formulas $\mathbf{SOL}(\bar{R})$ are obtained by allowing also relation variables $V_{i,\rho(i)}$, where $\rho(i)$ is the arity of $V_{i,\rho(i)}$, and atomic formulas of the type "$V_{i,\rho(i)}(x_1, \ldots x_{\rho(i)})$"

Monadic Second Order Logic formulas $\mathbf{MSOL}(\bar{R})$ are obtained by allowing as relation variables only set variables (unary predicates) U_i, atomic formulas of the type "$U_i(x_j)$" (also expressible as "$x_j \in U_i$"), and quantifications of the form $\exists U$ and $\forall U$.

Counting Monadic Second Order Logic formulas $\mathbf{CMSOL}(\bar{R})$ are obtained by allowing additional quantifiers for individual variables; for every $m, n \in \mathbb{N}$ we allow for the quantification "$C_{m,n}x$" – if $\phi(x)$ is a $\mathbf{CMSOL}(\bar{R})$-formula then so is $C_{m,n}x\phi(x)$.

The satisfaction relation between an $\bar{R}$-structure $\mathfrak{A}$ and an **SOL**-formula ϕ is defined as usual (for example, if there exists $A' \subset A$ such that $\mathfrak{A}$ satisfies $\phi(A')$, then $\mathfrak{A}$ satisfies $\exists U\phi(U)$), and is denoted by $\mathfrak{A} \models \phi$. For **CMSOL**, we define $\mathfrak{A} \models C_{m,n}x\phi(x)$ to hold if the number of elements $a \in A$ for which $\mathfrak{A} \models \phi(a)$ is equivalent to n modulo m.

A class of $\bar{R}$-structures $\mathcal{C}$ is is called $\mathbf{FOL}(\bar{R})$-*definable* if there exists an $\mathbf{FOL}(\bar{R})$ formula ϕ with no free (non-quantified) variables such that $\mathfrak{A} \in \mathcal{C}$ if and only if $\mathfrak{A} \models \phi$ for every $\mathfrak{A}$. We similarly define $\mathbf{MSOL}(\bar{R})$-definable classes and $\mathbf{CMSOL}(\bar{R})$-definable classes.

3.2. Proving non-definability. For the purpose of this article we just list a few useful facts and examples of classes of graphs definable and/or not definable in **CMSOL** and its sublogics.

First we note that for words, i.e., ordered structures with unary predicates only (besides the order relation), **MSOL** and **CMSOL** have the same expressive power. For graphs $G = (V, E)$, represented as structures with one binary edge relation E and a universe V of vertices, this is not the case.

Typical examples of classes of graphs not definable in **CMSOL** can therefore be obtained using reductions to non-regular languages, For example the language $a^n b^n$ over the alphabet $\{a, b\}$ is well known not to be regular, and therefore neither **MSOL**-definable nor **CMSOL**-definable. The class of complete bipartite graphs $K_{m,n}$ with a linear ordering is first order bi-reducible to the language $a^m b^n$. $K_{m,n}$ has a Hamiltonian cycle iff $m = n$, and therefore Hamiltonian graphs, even with a linear order, are not **CMSOL**-definable. The class $EQ_2CLIQUE$ of graphs consisting of two equal-sized cliques is also not **CMSOL**-definable, even in the presence of a linear order, because it consists of the complement graphs of $K_{n,n}$. Using the same method, one can construct other examples.

A typical example of a graph class which is not **MSOL**-definable, but **CMSOL**-definable, is the class of Eulerian graphs. A clique K_n is Eulerian iff n is odd. But the cliques of odd size are not **MSOL**-definable. An undirected graph G has an Eulerian cycle iff G is connected and every vertex has even degree. So graphs with Eulerian cycles are **CMSOL**-definable.

In contrast to this, the class of graphs where every induced cycle has an even size is **MSOL**-definable. To see this, one notes that an induced cycle is even iff it is bipartite.

4. Finiteness conditions

In order to prove modular linear recurrences (MC-finiteness) for a sequence $a(n)$ with an **MSOL**-interpretation $\mathcal{K}$ one proves first that $\mathcal{K}$ satisfies a certain finiteness condition derived using Ehrenfeucht-Fraïssé Games for **MSOL**. But often a weaker finiteness condition suffices to get the desired recurrence relation. We now discuss these finiteness conditions. They are all of the following form.

Let $\mathcal{K}$ be a class of τ-structures and let $\otimes$ be a binary operation on all τ-structures. We associate with $\mathcal{K}$ and $\otimes$ an equivalence relation $\sim_{\mathcal{K}}^{\otimes}$ on τ-structures: $\mathfrak{A}_1 \sim_{\mathcal{K}}^{\otimes} \mathfrak{A}_2$ iff for all τ-structures $\mathfrak{B}$ we have that $\mathfrak{A}_1 \otimes \mathfrak{B} \in \mathcal{K}$ iff $\mathfrak{A}_2 \otimes \mathfrak{B} \in \mathcal{K}$. Instead of $\mathfrak{A}_1 \sim_{\mathcal{K}}^{\otimes} \mathfrak{A}_2$ we also say that $\mathfrak{A}_1$ is $(\otimes, \mathcal{K})$-equivalent to $\mathfrak{A}_2$.

Our finiteness conditions require that the number of $\otimes(\mathcal{K})$-equivalence classes is finite. We call the number of $\otimes(\mathcal{K})$-equivalence classes the $\otimes$-index of $\mathcal{K}$ and denote it by $\otimes(\mathcal{K})$.

There are three operations $\otimes$ which are of interest:

(i) The *disjoint union* of τ-structures, denoted by $\sqcup$. In this case we speak of $(\sqcup, \mathcal{K})$-equivalence, and of the DU-index of $\mathcal{K}$, denote by $DU(\mathcal{K})$.
(ii) The *ordered sum* of τ_{ord}-structures, denoted by $\sqcup_{ord}$. In this case we speak of $(\sqcup_{ord}, \mathcal{K})$-equivalence, and of the OS-index of $\mathcal{K}$, denote by $OS(\mathcal{K})$.
(iii) The *substitution* of τ-structures into a pointed τ_a-structure, where τ_a has one distinguished constant symbol, and where the result is an unpointed

τ-structure. In this case we speak of $(subst, \mathcal{K})$-equivalence, and of the *Specker*-index of $\mathcal{K}$, denote by $SP(\mathcal{K})$.

(iv) Similarly, substitution is defined also for ordered structures, where all the elements of the substituted structure fall in their order between all the elements small than a and all the elements bigger than a, and the result is unpointed.

In the case of $\mathcal{K}$ being a set of words, the ordered sum corresponds to the concatenation of words and the OS-index is finite iff $\mathcal{K}$ is regular. This is the classical Myhill-Nerode characterization of regular languages. If we combine this with the Büchi-Elgot-Trakhtenbrot characterization of regular languages, we get that $\mathcal{K}$ has finite OS-index iff $\mathcal{K}$ is definable in **MSOL**.

In the case of $\mathcal{K}$ being a class of finite directed graphs, we say that $\mathcal{K}$ is a Gessel class, if $\mathcal{K}$ is closed under disjoint unions and components.

We shall prove in Section 13 that

THEOREM 4.1. *Let $\mathcal{K}$ be a class of τ-structures.*

(i) $DU(\mathcal{K}) \leq SP(\mathcal{K})$*;*
(ii) $OS(\mathcal{K}) \leq OSP(\mathcal{K})$*;*
(iii) If $\mathcal{K}$ is a Gessel class, then $DU(\mathcal{K}) \leq 2$.

Furthermore, we shall see in Section 14:

THEOREM 4.2. *If $\mathcal{K}$ is* **CMSOL***-definable then $SP(\mathcal{K})$ is finite.*

The theorems we discuss in this chapter all show that a sequence $a(n)$ satisfies some linear recurrence relation, provided $a(n)$ has a combinatorial interpretation with a finite $\otimes$-index for a suitable choice of $\otimes$.

To show that the $\otimes$-index is very small, say 2 or 3, a direct argument may suffice. However, to establish that the $\otimes$-index is finite, without computing an explict bound, it is often more convenient to use Theorem 4.2.

The exact relationship between finite indices and logical definability will be discussed in Section 14.

5. Logical interpretations of integer sequences

5.1. Logical interpretations of C-finite sequences. As our first example of the use of logical interpretations we reinterpret a classical theorem about regular languages. What we obtain is a characterization of C-finite sequences of integers as differences of two counting functions of regular languages.

Let $\mathcal{K}$ be a class of ordered structures with a fixed finite number of unary predicates. Such structures are conveniently identified with words over a fixed alphabet, and a class of such structures is called a language and is denoted by $L = \mathcal{K}$.

Let us recall the following characterization of languages which have finite OS-index, [**44, 26**].

THEOREM 5.1. *Let L be a language. The following are equivalent:*

(i) L has finite OS-index.
(ii) L is a regular language.
(iii) L is **MSOL***-definable.*

PROOF. The equivalence of (i) and (ii) is the classical Myhill-Nerode Theorem, whereas the equivalence of (ii) and (iii) was proven by Büchi, Trakhtenbrot and Elgot independently. □

THEOREM 5.2 (N. Chomsky and M. Schützenberger, [**18**]). *Let $d_L(n)$ be a counting function of a regular language L. Then $d_L(n)$ is C-finite.*

The converse is not true. However, we proved recently the following:

THEOREM 5.3 ([**45**]). *Let $a(n)$ be a function $a : \mathbb{N} \rightarrow \mathbb{Z}$ which is C-finite. Then there are two regular languages L_1, L_2 with counting functions $d_1(n), d_2(n)$ such that $a(n) = d_1(n) - d_2(n)$.*

The proof will be given in Section 5.1.

REMARK 5.1. *We could replace the difference of two sequences in the expression $a(n) = d_1(n) - d_2(n)$ by $a(n) = d_3(n) - c^n$ where $d_3(n)$ also comes from a regular language, and $c \in \mathbb{N}$ is suitably chosen.*

Using the characterization of regular languages of Theorem 5.1, Theorem 5.2 and Theorem 5.3 can be combined to characterize the C-finite sequences of integers.

THEOREM 5.4. *Let $a(n)$ be a function $a : \mathbb{N} \rightarrow \mathbb{Z}$. $a(n)$ is C-finite iff there are two* **MSOL**-*Specker sequences $d_1(n), d_2(n)$, where the sequences $d_1(n), d_2(n)$ have* **MSOL**-*interpretations over a fixed finite vocabulary which contains $<_{nat}$ and otherwise only unary relation symbols, such that $a(n) = d_1(n) - d_2(n)$.*

5.2. Logical interpretations of P-recursive (holonomic) sequences. An infinite set of holonomic sequences can be obtained from counting restricted lattice walks. A step in a lattice walk is a pair $a = (x, y) \in \mathbb{Z}^2$. For a set $\mathcal{Y}$ of steps, a lattice walk is a word $w \in \mathcal{Y}^\star$. $\mathcal{Y}$ is *symmetric* if for all $(i, j) \in \mathcal{Y}$ also $(i, -j) \in \mathcal{Y}$. For a lattice walk $w = (x_1, y_1)(x_2, y_2) \ldots (x_m, y_m)$ we define $X_i(w) = \sum_{j \leq i} x_j$ and $Y_i(w) = \sum_{j \leq i} y_j$. A lattice walk over the *quarter plane* is a lattice walk w such that for all $i \leq \ell(w)$ we have that $X_i(w), Y_i(w) \geq 0$. A lattice walk over the quarter plane stays below the diagonal, if additionally we have for all $i \leq \ell(w)$ that $Y_i(w) \leq X_i(w)$. An *n-lattice path* is a self-avoiding lattice walk starting at a point $(0, y_0)$ and ending at a point (n, Y_m). Usually one counts lattice walks by their length m and lattice paths by prescribing their origin and their end points on an $(n \times n)$-grid. Note that an n-path may have length different from n.

The next two theorems show that there are different ways of counting lattice walks and paths (or combinations thereof) which yield holonomic sequences. In analogy to Theorems 5.3 and 5.4, we are looking for a way to represent all holonomic sequences in a uniform way.

Let $\mathcal{Y} \subset \mathbb{Z} \times \mathbb{Z} - \{(0, 0)\}$.

Denote by $A_{\mathcal{Y},q}(m)$ ($A_{\mathcal{Y},d}(m)$) the number of lattice walks of length m, i.e., words of length m, in the quarter plane with steps in $\mathcal{Y}$ (which stay below the diagonal).

Denote by $a_{\mathcal{Y},q}(n)$ ($a_{\mathcal{Y},d}(n)$) be the number of n-lattice paths in the quarter plane with steps in $\mathcal{Y}$ (which stay below the diagonal). For example, if $\mathcal{Y} = \{1\} \times \mathbb{Z}$, then $a_{\mathcal{Y},q}(n)$ is infinite, but $a_{\mathcal{Y},d}(n) = n!$.

THEOREM 5.5 (Bousquet-Mélou, [**16**]). *Let $\mathcal{Y} \subset \{-1, 0, 1\} \times \mathbb{Z} - \{(0, 0)\}$. If $\mathcal{Y}$ is finite and symmetric, then $A_{\mathcal{Y},q}(m)$ is holonomic.*

The symmetry assumption on $\mathcal{Y}$ is essential. In [**54**] it is shown that there are asymmetric finite $\mathcal{Y} \subseteq \{1,-1,0\}^2 - \{(0,0)\}$ such that $A_{\mathcal{Y},q}(m)$ is not holonomic.

To get many examples, we make the set of allowed steps dependent on its position. Let $w \in \mathcal{Y}^n$, $a \in \Sigma$ and $u \in \Sigma^n$. We say that w follows u at a if the following holds: If $u[i] = a$ then $w[i] \in \mathcal{Y}^n$, else $w[i] = (1,0)$.

$a_{\mathcal{Y},d,L,a}(n)$ counts the number of pairs (w,u) such that $u \in L$ and $\ell(w) = \ell(u) = n$, and w is an n-path below the diagonal which follows u at a. Similarly, for $\bar{a} = (a_1, \ldots, a_k)$ the function $a_{\mathcal{Y},d,L,\bar{a}}(n)$ counts the number of tuples $(w_1, \ldots w_k, u)$ such that $u \in L$ and for $j \leq k$, we have that $\ell(w_j) = \ell(u) = n$ and w_j is a path below the diagonal which follows u at a_j.

Definition 5.1. *Let $\mathcal{Y} = \{1\} \times \{-1,0,1\}$. A sequence $d(n)$ of integers is an LP-sequence if there is a regular language $L \subset \Sigma^\star$ and elements $a_1, \ldots, a_k \in \Sigma$ such that $d(n) = a_{\mathcal{Y},d,L,\bar{a}}(n)$.*

Theorem 5.6 ([**46**]). *Let $d(n)$ be an LP-sequence of integers. Then $d(n)$ is holonomic, and even SP-recursive.*

LP-sequences count combinations of lattice paths with a fixed set $\mathcal{Y}$ of steps, and which all follow along words of length n of a suitable regular language L. A further degree of freedom is given by the choice of letters $\bar{a}$.

Conversely, every SP-recursive sequence can be obtained from sequences with LP-interpretations:

Theorem 5.7 ([**46**]). *A sequence $d(n)$ of integers is SP-recursive iff there are two LP-sequences $d_1(n), d_2(n)$ such that $d(n) = d_1(n) - d_2(n)$.*

We consider Theorem 5.7 a "logical characterization" of SP-recursive sequences in the same sense as Theorems 5.3 and 5.4 are a "logical characterization" of C-finite sequences. The general case of characterizing P-recursive (holonomic) sequences will be discussed in Section 12. In Subsection 12.4 we discuss a different approach for characterizing holonomic sequences which uses position specific weights on words from [**47**].

5.3. Logical interpretations and modular recurrences. Modular recurrence relations for sequences with combinatorial interpretations are studied widely, cf. [**33, 38**]. A logical approach to this topic was pioneered in [**13, 14**] and further pursued in [**66, 67**]. In this section we only outline what we discuss in greater detail in Part 4.

C. Blatter and E. Specker have shown:

Theorem 5.8 (C. Blatter and E. Specker, [**13**]). *Let $a(n)$ be a Specker sequence which has a* pure *combinatorial interpretation $\mathcal{K}$ over a finite vocabulary which contains only relation symbols of arity at most 2. If $\mathcal{K}$ has finite Specker index then $a(n)$ is MC-finite.*

Using Theorem 4.2 we get

Corollary 5.1. *Let $a(n)$ be a Specker sequence which has a* pure *combinatorial interpretation $\mathcal{K}$ over a finite vocabulary which contains only relation symbols of arity at most 2. If $\mathcal{K}$ is* **MSOL**-*definable then $a(n)$ is MC-finite.*

Remarks 5.1.

(i) *Corollary 5.1 is not true for* **MSOL***-interpretations with an order, i.e. which are not pure, cf.* [**32**].
(ii) *E. Fischer,* [**31**]*, showed that Corollary 5.1 is also not true if one allows relation symbols of arity* ≥ 4*, see also* [**67**].
(iii) *In the light of Proposition 1.2(i) and Proposition 2.3(ii) there cannot be a converse of Corollary 5.1, the set of MC-finite sequences of integers cannot be characterized by a set of Specker sequences.*

In 1984 I. Gessel proved the following related result:

THEOREM 5.9 (I. Gessel, [**38**]). *Let* $\mathcal{K}$ *be a class of (possibly) directed graphs of bounded degree* d *which is a Gessel class, i.e., closed under disjoint unions and components. Then* $d_{\mathcal{K}}(n)$ *is MC-finite.*

REMARK 5.2. *Theorem 5.9 does not use logic. However, let* $\mathcal{K}$ *be a class of connected finite directed graphs, and let* $\bar{\mathcal{K}}$ *be the closure of* $\mathcal{K}$ *under disjoint unions. It is easy to see that* $\mathcal{K}$ *is* **MSOL***-definable iff* $\bar{\mathcal{K}}$ *is* **MSOL***-definable. Therefore, naturally arising Gessel classes are likely to be definable in* **SOL** *or even* **MSOL**.

The notion of degree can be extended to arbitrary relational structures $\mathcal{A}$ via the Gaifman graph of $\mathcal{A}$, cf. [**25**].

DEFINITION 5.1.
(i) *Given a structure* $\mathfrak{A} = \langle A, R_1^A, \ldots, R_k^A \rangle$*,* $u \in A$ *is called a* neighbor *of* $v \in A$ *if there exists a relation* R_i^A *and some* $\bar{a} \in R_i^A$ *containing both* u *and* v.
(ii) *We define the* Gaifman graph $Gaif(\mathfrak{A})$ *of a structure* $\mathfrak{A}$ *as the graph with the vertex set* A *and the neighbor relation defined above.*
(iii) *The* degree *of a vertex* $v \in A$ *in* $\mathfrak{A}$ *is the number of its neighbors. The degree of* $\mathfrak{A}$ *is defined as the maximum over the degrees of its vertices. It is the degree of its Gaifman graph* $Gaif(\mathfrak{A})$.
(iv) *A structure* $\mathfrak{A}$ *is* connected *if its Gaifman graph* $Gaif(\mathfrak{A})$ *is connected.*

Inspired by Theorem 5.8 and Theorem 5.9, E. Fischer and J.A. Makowsky showed:

THEOREM 5.10 (E. Fischer and J.A. Makowsky, [**32**]). *Let* $a(n)$ *be a Specker sequence which has a* pure *combinatorial-interpretation* $\mathcal{K}$ *over any finite relational vocabulary (without restrictions on the arity of the relation symbols), but which is of bounded degree.*
(i) *If* $\mathcal{K}$ *has finite DU-index then* $a(n)$ *is MC-finite.*
(ii) *If additionally all structures in* $\mathcal{K}$ *are connected, then* $a(n)$ *is trivially MC-finite.*

The proof will be given in Section 15.

Part 2. Guiding Examples

We now discuss various combinatorial functions with respect to their logical interpretations and the nature of their recurrence relations.

6. The classical recurrence relations

6.1. C-finite with positive coefficients. The Fibonacci sequence $f(n)$ is defined by $f(n+2) = f(n+1) + f(n)$ with $f(1) = 1$ and $f(2) = 2$. It is therefore C-finite.

To illustrate Theorem 5.3, let L_{Fib} be given by the regular expression $(a \vee ab)^\star$ with counting function $d_{Fib}(n)$. It is easy to see that $d_{Fib}(n) = f(n)$. Similarly, if $g(n+2) = 2g(n+1) + 3g(n)$ and L_g is given by

$$(a_1 \vee a_2 \vee b_1^2 \vee b_2^2 \vee b_3^2)^\star$$

with counting $d_g(n)$, then $g(n) = d_g(n)$.

It is not difficult to generalize this to any C-finite sequence with positive coefficients. For the general case, see [**7, 45**].

6.2. Growth arguments. Let $\mathcal{K}$ be a class of graphs and $d_{\mathcal{K}}(n)$ its counting function. We denote by $\bar{\mathcal{K}}$ the complement of $\mathcal{K}$. If $\mathcal{K}$ is the class of all graphs, then $d_{\mathcal{K}}(n)$ is not holonomic, since $2^{\binom{n}{2}}$ grows faster than $(n!)^c$. It follows that for any graph property $\mathcal{K}$ either $d_{\mathcal{K}}(n)$ or $d_{\bar{\mathcal{K}}}(n)$ is not holonomic, because the sum of two holonomic sequences is again holonomic.

Let $\mu_{\mathcal{K}}(n) = \frac{d_{\mathcal{K}}(n)}{2^{\binom{n}{2}}}$ denote that fraction of graphs of size n which are in $\mathcal{K}$. From the above we immediately get:

Proposition 6.1. *Let $\mathcal{K}$ be any graph property, i.e., a class of finite simple graphs closed under isomorphisms.*

(i) If $\lim \mu_{\mathcal{K}}(n) = \alpha$ exists and $\alpha \neq 0$, then $d_{\mathcal{K}}(n)$ is not holonomic.
(ii) If $\lim \mu_{\mathcal{K}}(n) = \alpha$ exists and $\alpha < 1$ then $d_{\bar{\mathcal{K}}}(n)$ is not holonomic.

7. Functions, Permutations and Partitions

7.1. Factorials. The factorial function $n!$ is SP-recursive and hypergeometric by $(n+1)! = (n+1) \cdot n!$. This shows that it is also trivially MC-finite. $n!$ is not C-finite because it grows too fast.

$n!$ has several combinatorial interpretations: It counts the number of functions $f : [n] \to [n]$ which are bijective, which is pure and **MSOL**-definable, and it also counts the number of functions $f : [n] \to [n]$ such that $f(i) <_{nat} i+1$, which is not pure, but also **MSOL**-definable.

7.2. The function n^n. The function n^n is not P-recursive, [**36**]. It is MC-finite, which is an easy consequence of the Little Fermat Theorem, i.e., $n^{p-1} = 1 \pmod p$ provided p does not divide n. The function n^n counts the number of structures on universe $[n]$ over vocabulary $\tau_F = \langle F \rangle$, where F is an unary function symbol. In other words, n^n is the number of functions $f : [n] \to [n]$. This gives a pure **MSOL**-definable interpretation.

7.3. Derangement numbers. The derangement numbers $D(n)$ are usually defined by their pure combinatorial definition as the set of functions $f : [n] \to [n]$ such that f is bijective and for all $i \in [n]$ we have $f(i) \neq i$. This is **MSOL**-definable. Their explicit definition is given by

$$D(n) = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!} = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor$$

They are SP-recursive by $D(n) = (n-1)(D(n-1) + D(n-2))$ with $D(0) = 1$ and $D(1) = 0$, hence MC-finite, but not C-finite, by the growth argument from Subsection 6.2. In [**57**, Example 8.6.1.] they are shown not to be hypergeometric.

7.4. Central binomial coefficient. The function $\binom{2n}{n}$, central binomial coefficient, is P-recursive and hypergeometric by

$$(n+1)^2 \cdot \binom{2(n+1)}{n+1} = 2 \cdot \binom{2(n+1)}{2} \cdot \binom{2n}{n}$$

$\binom{2n}{n}$ has many combinatorial interpretations: It counts the number of ordered partitions of $[2n]$ into two equal sized sets. If the partitions are not ordered, the counting functions is $\frac{1}{2}\binom{2n}{n}$.

$\binom{2n}{n}$ also counts the number of functions $f : [n+1] \to [n+1]$ such that $f(i+1) \geq f(i)$ and $f(n+1) = n+1$. This is not pure but **MSOL**-definable. $\frac{1}{2}\binom{2n}{n}$ also counts the number of graphs with vertex set $[2n]$ which consists of the disjoint union of two equal sized cliques. We denote this class of graphs by $EQ_2CLIQUE$. Both of the above combinatorial interpretations are pure, but not **MSOL**-definable in the language of graphs, [**66**].

Similarly, the class $EQ_pCLIQUE$ denotes the class of graphs with vertex set $[pn]$ which consist of p disjoint cliques of equal size. We denote by $b_p(n)$ the number of graphs with $[n]$ as a set of vertices which are in $EQ_pCLIQUE$. Clearly,

$$b_p(n) = \begin{cases} \frac{1}{p!}\binom{pn}{n} \cdot \binom{(p-1)n}{n} \cdot \ldots \cdot \binom{n}{n} & \text{if } p \text{ divides } n \\ 0 & \text{else} \end{cases}$$

Congruence relations of binomial coefficients and related functions have received a lot of attention in the literature, starting with Lucas's famous result for $b_2(n)$, [**50**]. For $b_p(n)$ modulo p, a prime, we have:

LEMMA 7.1. *For every $k > 1$, $b_p(pk) \equiv b_p(k) \pmod p$.*

The proof uses the method of combinatorial proof of Fermat's congruence theorem by J. Petersen from 1872, given in [**38**, page 157].

PROPOSITION 7.2. *For every n which is not a power of the prime p, we have $b_p(n) \equiv 0 \pmod p$, and for every n which is a power of p we have $b_p(n) \equiv 1 \pmod p$. In particular, $b_p(n)$ is not ultimately periodic modulo p.*

PROOF. By induction on n, where the basis is $n = p$ (for which $b_p(n) = 1$) and every n which is not divisible by p (for which $b_p(n) = 0$); the induction step follows from Lemma 7.1. □

From the above one can see that $b_2(n) = \frac{1}{2}\binom{2n}{n}$, and more generally $b_p(n)$, are P-recursive, but are not MC-finite. Therefore, for p a prime, $b_p(n)$ is is neither SP-recursive nor C-finite.

7.5. Bell numbers. The Bell numbers B_n count the number of partitions of an n-element set. They also count the number of equivalence relations on an n-element set, which gives a pure **MSOL**-interpretation, and Theorem 5.8 applies, hence they are MC-finite. Theorem 5.8 only proves the existence of modular recurrence relations, and no explicit modular recurrence relations appear in the literature which cover all $m \in \mathbb{N}$. For prime moduli p, however, they satisfy the simple Touchard congruence $B_{p+n} = B_n + B_{n+1} \pmod{p}$, [**37**]. They satisfy the recurrence relation

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k,$$

but in [**36**] it is shown that they are not holonomic. For more properties of Bell Numbers, cf. [**61**] and for congruences, cf. [**37**].

7.6. Stirling numbers of the first kind. The Stirling numbers of the first kind $\left[{n \atop k}\right]$ count arrangements of $[n]$ into k non-empty cycles (where a single element and a pair of elements are considered cycles). In other words, $\left[{n \atop k}\right]$ counts permutations with k cycles. For fixed k this is a pure **MSOL**-interpretation. They satisfy the following recurrence relation

$$\left[{n \atop k}\right] = (n-1)\left[{n-1 \atop k}\right] + \left[{n-1 \atop k-1}\right] .$$

Using the growth argument from Subsection 6.2 we can see that the Stirling numbers of the first kind are not C-finite, because $\left[{n \atop 1}\right]$ grows like the factorial $(n-1)!$. Using the Cayley-Hamilton theorem, one can deduce from the above equation that for fixed k, the sequence $\left[{n \atop k}\right]$ is SP-recursive, and therefore also MC-finite.

7.7. Stirling numbers of the second kind. The Stirling numbers of the second kind $\left\{{n \atop k}\right\}$ count the number of partitions of $[n]$ into k non-empty parts. For fixed k this is a pure **MSOL**-interpretation. $\left\{{n \atop k}\right\}$ is given explicitely by sum

$$\left\{{n \atop k}\right\} = \frac{1}{k!}\sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i} i^n .$$

They satisfy the following recurrence relation

$$\left\{{n \atop k}\right\} = k\left\{{n-1 \atop k}\right\} + \left\{{n-1 \atop k-1}\right\} .$$

Writing this for all $k' \leq k$, we can extract from this recurrence, using the Cayley-Hamilton Theorem, that for fixed k the sequence $\left\{{n \atop k}\right\}$ is C-finite. We shall show (Proposition 11.1) in Section 5.1 that $\left\{{n \atop k}\right\}$ is C-finite without using the above recurrence relation. Instead we will exhibit a regular language the counting function of which is given by the Stirling numbers of the second kind.

Note that the Bell numbers can be expressed in terms of the Stirling numbers of the second kind:

$$B_n = \sum_{i=0}^{n} \left\{{n \atop i}\right\} .$$

7.8. Catalan numbers. The Catalan numbers $C(n)$ are defined by $C(n) = \frac{1}{n+1} \cdot \binom{2n}{n}$. They satisfy the recurrence relation

$$C(n+1) = \frac{2(2n+1)}{n+2} \cdot C(n) \tag{1}$$

In general, concerning the recurrence relations they behave like $\frac{1}{2}\binom{2n}{n}$. In [**69**] there is an abundance of combinatorial interpretations which are not pure. Many of these are based on functions $f : [n] \to [n]$ which represent lattice paths subject to various conditions. One of these is the set of weakly monotonic functions $f : [n] \to [n]$ such that $f(1) = 1$, $f(n) = n$ and $f(i) \leq i$.

8. Trees and Forests

8.1. Trees. Trees are (undirected) connected acyclic graphs. They are not **FOL**-definable but are **MSOL**-definable, and have finite Specker index, hence finite DU-index. Acyclicity is expressed by saying that there is no subset of size at least three such that the induced graph on it is 2-regular and connected. Denote by T_n the number of of labeled trees on n vertices. From the **MSOL**-definability it follows that the sequence T_n is MC-finite.

Labeled trees were among the first objects to be counted explicitely, cf.[**41**, Theorem 1.7.2].

THEOREM 8.1 (A. Cayley 1889). *$T_n = n^{n-2}$.*

Here the modular linear recurrences can be given explicitly: For $m = 2$ we have

$$T_1 = T_2 = 1, T_3 = 3, T_4 = 16, T_5 = 125, \dots$$

and $T_n = n \pmod 2$ for $n \geq 3$. The function n^{n-2} is not P-recursive, [**36**]. The same holds for n^{n-1} which counts the number of rooted trees.

For the number of trees of outdegree bounded by k we get the following corollary of Theorem 5.10:

COROLLARY 8.2. *The number of labeled trees of outdegree at most k is, for every $m \in \mathbb{N}$, ultimately constant $\pmod m$.*

In [**41**, Chapter 3] there is a wealth of results on counting various labeled trees and tree-like structures. It is worth noting that the notion of k-tree, and more generally the property of a graph of having tree-width at most k are **MSOL**-definable, cf. [**22**].

8.2. Forests. Forests are disjoint unions of trees, or equivalently, they are acyclic graphs. Therefore they are **MSOL**-definable, and have finite Specker index. In fact, they form a Gessel class.

It is well known, cf. [**73**], that the number of rooted forests on n vertices is $RF_n = (n+1)^{n-1}$. Again this is not holonomic but MC-finite.

The number of forests F_n (of non-rooted trees) is more complicated. L. Takács, [**72**] showed that

$$F_n = \frac{n!}{n+1} \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^j \frac{(2j+1)(n+1)^{n-2j}}{2^j \cdot j! \cdot (n-2j)!} \tag{2}$$

A simpler proof is given in [**17**].

From Theorem 5.8 we know that F_n is MC-finite, which is not obvious from the formula, as it is not obvious that the sum has integer value. Whether F_n is holonomic or not seems to be open.

9. Graph properties

In this section we list examples of graph properties $\mathcal{K}$. By definition $d_{\mathcal{K}}(n)$ has a pure combinatorial interpretation. We discuss the definability of $\mathcal{K}$ and the behaviour of $d_{\mathcal{K}}(n)$ in terms of recurrence relations. Our main sources for definability are [**25, 22**], and for the behaviour of $d_{\mathcal{K}}(n)$, [**41, 73**].

9.1. Connected graphs. The class $\mathcal{K}$ = CONN is not **FOL**(R)-definable, but it is **MSOL**(R)-definable using a universal quantifier over set variables. We just say that every subset of vertices which is closed under the edge relation has to be the set of all vertices. For a detailed discussion of the exact status of definability, cf. [**2**].

Counting labeled connected graphs is treated in [**41**, Chapters 1 and 7] and in [**73**, Chapter 3]. For CONN [**41**, page 7] gives the following recurrence:

$$d_{\text{CONN}}(n) = 2^{\binom{n}{2}} - \frac{1}{n}\sum_{k=1}^{n-1} k\binom{n}{k} 2^{\binom{n-k}{2}} d_{\text{CONN}}(k).$$

It is well known, see [**49**, Page 236], that $\lim_{n\to\infty}\mu_{\text{CONN}}(n) = 1$, To see that $d_{\text{CONN}}(n)$ is not holonomic we use Proposition 6.1.

9.2. Regular graphs. The class REG_r of simple regular graphs where every vertex has degree r is **FOL**-definable (for fixed r). The formula says that every vertex has exactly r different neighbors. The formula grows with r. The class REG of regular graphs without specifying the degree is not **FOL**-definable, and actually not even **CMSOL**-definable. To see this we note that a complete bipartite graph $K_{m,n}$ is regular iff $m = n$, but equi-cardinality of definable relations is not **CMSOL**-definable. The class CREG_r of connected r-regular graphs is **MSOL**-definable (for fixed r).

Counting the number of labeled regular graphs is treated completely in [**41**, Chapter 7], where an explicit formula is given, essentially due to J.H. Redfield [**60**] and rediscovered by R.C. Read [**58, 59**]. However, the formula is very complicated.

For cubic graphs, the function is explicitly given in [**41**, page 175] as $d_{\mathcal{R}_3}(2n+1) = 0$ and

$$d_{\mathcal{R}_3}(2n) = \frac{(2n)!}{6^n}\sum_{j,k}\frac{(-1)^j(6k-2j)!6^j}{(3k-j)!(2k-j)!(n-k)!}48^k\sum_i\frac{(-1)^i j!}{(j-2i)!i!}$$

Both REG and REG_r are Gessel classes, i.e., closed under taking compononts and disjoint unions. Furthermore, REG_r is of bounded degree. Applying Theorem 5.9, we see that $d_{\text{REG}_r}(n)$ is MC-finite with a simple two term recurrence relation. Applying Theorem 5.10 we get that $d_{\text{CREG}_r}(n)$ is trivially MC-finite.

I. Gessel [**40**] showed that for fixed $r \in \mathbb{N}$ the sequence $d_{\text{REG}_r}(n)$ is holonomic. The problem of counting r-regular graphs with a specified set of forbidden subgraphs is one whose holonomicity remains open. N.C. Wormald [**74**] showed that the counting sequence for 3-regular graphs without triangles is holonomic.

Let h_n be the number of claw-free cubic graphs on $2n$ labeled nodes. Recently, E. Palmer, R. Read and R. Robinson, [**56**], have shown that "the enumeration of labeled claw-free cubic graphs can be added to the handful of known counting problems for regular graphs with restrictions which have been proved P-recursive". Actually, they showed that h_n is SP-recursive, giving explicitly a linear homogeneous recurrence of order 12 in which the coefficients are polynomials of degree up to 23 and all have integer coefficients. Therefore, h_n is also MC-finite. However, MC-finiteness for all the above examples follows also, without giving the recurrence explicitly, since the classes of triangle-free, or claw-free cubic graphs are **FOL**-definable.

9.3. Bipartite graphs. The class BIPART of bipartite graphs is **MSOL**-definable, and so is the class of connected bipartite graphs. We say that there is partition of the vertex set into two independent sets (and add the statement for connectedness). Let $\beta(n)$ be the number of labeled bipartite graphs. Therefore $\beta(n)$ is MC-finite. Furthermore, the counting function for r-regular connected bipartite graphs is trivially MC-finite. BIPART is also a Gessel class. Therefore, the counting function for r-regular bipartite graphs is MC-finite with a simple two term recurrence relation.

Note that the class BIPART is not **FOL**-definable, since a regular graph of degree two is bipartite iff all its components are cycles of even size. But large enough cycles of even or odd degree cannot be distinguished by an **FOL**-formula.

In [**41**, Page 17] an explicit formula is given:

$$\beta(n) = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k}2^{k(n-k)}$$

This also shows that $\beta(n)$ is not holonomic, because it grows too fast.

9.4. Graphs of even degree and Eulerian graphs. Let $\mathcal{C} = EVENDEG$ be the class of simple graphs where each vertex has even degree. $EVENDEG$ is not **MSOL**-definable but is **CMSOL**-definable. By Theorem 4.2 it has finite Specker index, hence finite DU-index.

$d_{EVENDEG}(n) = 2^{\binom{n-1}{2}}$, cf. [**41**, page 11].

This is an example where the same function has two combinatorial interpretations: $d_{EVENDEG}(n+1) = d_{GRAPHS}(n)$ where the former is not **MSOL**-definable, but the latter is even **FOL**-definable.

Let $\mathcal{C} = EULER$ be the class of simple connected graphs in $EVENDEG$. $EULER$ is not **MSOL**-definable, but is **CMSOL**-definable. In [**41**, page 7] the following recurrence for $d_{\mathrm{EULER}}(n)$ is given:

$$d_{\mathrm{EULER}}(n) = 2^{\binom{n-1}{2}} - \frac{1}{n}\sum_{k=1}^{n-1} k\binom{n}{k}2^{\binom{n-k-1}{2}}d_{\mathrm{EULER}}(k).$$

The number of Eulerian graphs of degree at most r is also **CMSOL**-definable. To find an explicit formula of its counting function seems very hard. However, our Theorem 5.10 gives that the number of such graphs is trivially MC-finite.

9.5. Planar graphs and grid graphs. Planar graphs are **MSOL**-definable. To see this one can use the Kuratowski-Wagner Theorem characterizing planar graphs with forbidden (topological) minors, cf. [**24**].

A special kind of planar graphs are the *rectangular grids* $GRIDS$, which look like rectangular checker boards, with the north-south and east-west neighborhood relation. *Partial rectangular grids* $PGRIDS$ are subgraphs of rectangular grids. It is easy to see that both $GRIDS$ and $PGRIDS$ have finite Specker index, but $GRIDS$ are **MSOL**-definable while $PGRIDS$ are not **CMSOL**-definable, cf. [**20, 22, 65, 62**].

9.6. Forbidden subgraphs and forbidden minors. To get many classes of graphs which are **MSOL**-definable it is useful to note the following:

PROPOSITION 9.1. *Let $\mathcal{H}$ be a fixed finite set of finite graphs. Then the following are* **MSOL**-*definable:*

(i) *The class of graphs which have no subgraph isomorphic to some $H \in \mathcal{H}$.*
(ii) *The class of graphs without an induced subgraph isomorphic to some $H \in \mathcal{H}$.*
(iii) *The class of graphs without a minor isomorphic to some $H \in \mathcal{H}$.*
(iv) *The class of graphs without a topological minor isomorphic to some $H \in \mathcal{H}$.*

Proof. (i) and (ii) are easily expressed on **FOL**.

We sketch a proof of (iii) in the case where $\mathcal{H}$ consists of a single graph. Assume $H = (V(H), E(H))$ with $V(H) = \{1, \ldots, , n\}$. Now G has H as a minor if and only if the following holds:

there are disjoint connected subsets $U_i \subseteq V(G)$, $i = 1, \ldots, n$,
such that there is an edge $e \in E(G) \cap (U_i \times U_j)$ if $(i, j) \in E(H)$.

Clearly this can be expressed in **MSOL**.

To prove (iv), we note that H is a topological minor of G if and only if for each edge e of $E(H)$, there is a set U_e of vertices inducing a connected subgraph such that

(i) if e_1, e_2, e_3 are distinct edges of $E(H)$, then $U_{e_1} \cap U_{e_2} \cap U_{e_3} = \emptyset$,
(ii) if e_1, e_2 are distinct edges of $E(H)$, then $U_{e_1} \cap U_{e_2} \neq \emptyset$ iff e_1 and e_2 have a common vertex, and
(iii) if $U_{e_1} \cap U_{e_2} \neq \emptyset$ then it has exactly one element.

Again these conditions can be expressed in **MSOL**.

□

It follows that every minor closed class of graphs is **MSOL**-definable. To see this, one uses the Graph Minor Theorem of P. Robertson and N. Seymour, which states that every minor closed class of graphs can be represented as a class of graphs with a finite set of forbidden minors, see [**24**].

The power of **MSOL** in graph theory has been studied extensively by B. Courcelle and J. Engelfriet [**23**].

9.7. Perfect graphs. A graph is perfect if for every induced subgraph (including the graph itself) the chromatic number equals the clique number. On the face of it, this does not seem **MSOL**- or **CMSOL**-definable. However, the following was conjectured by Berge [**15**, Chapter V.5] and proved by M. Chudnovsky, N. Robertson, P. Seymour, R. Thomas in [**19**].

Theorem 9.1 (Strong Perfect Graph Theorem). *A graph G is perfect iff neither G nor its complement graph contains an odd induced cycle of size at least* 5.

This gives us an **MSOL**-definition of perfect graphs. Furthermore, the Specker index for perfect graphs can be computed directly and is much smaller than one would get using the **MSOL**-definition.

Proposition 9.2. *If G and H are graphs, and a is a vertex of G, then $Subst(G, a, H)$ is perfect iff both G and H are perfect.*

Proof. One direction follows from the definition, the other direction is by now classic, cf. [**15**, Chapter V.5, Theorem 19]. □

Using Corollary 13.4 we get

Corollary 9.3. *The Specker index of perfect graphs is* 2.

9.8. More CMSOL-definable classes. We can now combine previous properties and see that the following are **CMSOL** definable classes of graphs. We have not found any explicit discussion of their counting functions in the literature, but the Specker-Blatter Theorem (or one of our generalizations) applies to all of these cases. The following list can be extended ad libitum.

- Planar Eulerian graphs and Eulerian graphs of .bounded degree d.
- Graphs where every clique has odd cardinality.
- Graphs where every minimal cycle has even cardinality.
- Planar regular graphs of odd (even) degree.
- Planar graphs with a finite set of forbidden induced subgraphs.

10. Latin Squares

In the Specker-Blatter Theorem, 5.8, the pure combinatorial interpretations are required to use relations of arity at most 2. The theorem is known to be false with arity 4, as was shown by E. Fischer, [**31**], and its status is open for arity 3.

Latin rectangles are matrices of size $(k \times n)$ with entries from $[n]$ such that in each row and column each number appears at most once. Latin squares are of the form $(n \times n)$. We denote by $L(k, n)$ the number of Latin rectangles of size $(k \times n)$ A Latin rectangle is reduced if the the first row is $(1, 2, \dots, n)$ and the first column is $(1, 2, \dots, k)$. We denote by $R(k, n)$ the number of reduced Latin rectangles of size $(k \times n)$. It is well known, [**52**], that

$$L(k, n) = \frac{n!(n-1)!}{(n-k)!} \cdot R(k, n) \tag{3}$$

The sequences $L(n, n)$ and $n \cdot R(n, n)$ have pure **MSOL**-interpretations with one ternary predicate. To see this we note that $L(n, n)$ counts the number of relations $M \subseteq [n]^3$ such that

- for every $i, j \in [n]$ there is exactly one $k \in [n]$ with $(i, j, k) \in M$, and
- for every $i, k \in [n]$ there is exactly one $j \in [n]$ with $(i, j, k) \in M$, and
- for every $k, j \in [n]$ there is exactly one $i \in [n]$ with $(i, j, k) \in M$.

Similarly, $n \cdot R(n, n)$ counts the number of relations $M \subseteq [n]^3$ which additionally satisfy

- there is $i \in [n]$ such that for all $j \in [n]$ we have $(i, j, j) \in M$, and $(j, i, j) \in M$.

For fixed k the sequences $L(k, n)$ and $n \cdot R(k, n)$ have pure **MSOL**-interpretations with k binary predicates $P_i(j, k)$ and the corresponding properties.

From Equation (3) it follows that $L(k, n)$ and $L(n, n)$ are both trivially MC-finite. This is also true for $R(k, n)$ or $R(n, n)$, by a theorem due to E.B. McKay and I.M. Wanless [**53**], cf. also [**70**, Theorem 4.1.], as they proved:

THEOREM 10.1. *Let $m = \lfloor n/2 \rfloor$. For all $n \in N$, $R(n, n)$ is divisible by $m!$.*

D. S. Stones and I.M. Wanless, [**71**] also showed, cf. [**70**, Theorem 4.4.]:

THEOREM 10.2. *If $k \leq n$ then $R(k, n+t) = ((-1)^{k-1}(k-1)!)^t \cdot R(k, n) \pmod t$ for all $t \geq 1$.*

In some cases, this shows that $R(k, n)$ is indivisible by some t for all $n > k$, when k is fixed and $t > k$. Nevertheless, Theorem 5.8 shows that, for fixed k, the sequence $R(k, n)$ is MC-finite.

On the other hand, $L(n,n)$ is not holonomic, as by [**27**] $L(k,n)$ grows asymptotically like

$$L(k,n) \approx (n!)^k \cdot exp\left(\frac{-k(k-1)}{2}\right). \tag{4}$$

Using Equation (3), it follows that $R(n,n)$ is also not holonomic. For fixed k, I. Gessel has shown, [**39**], that $R(k,n)$ and $L(k,n)$ are holonomic, without giving the recurrence explicitly. From Equation (4) one can also see that they are not C-finite.

Part 3. C-Finite and Holonomic Sequences

11. C-Finite sequences

We now return to the characterization of C-finite sequences as stated in Section 5. In Subsection 11.1 we prove the missing direction of Theorem 5.3 as follows:

THEOREM 11.1. *Let $a(n)$ be a C-finite sequence of integers. There exist a regular language L and a constant $c \in \mathbb{N}$ such that $a(n) = d_L(n) - c^n$, where $d_L(n)$ is the number of words of length n in L.*

The proof is based on the proof presented in [**45**], but uses the framework of the theory of regular languages instead of Monadic Second Order Logic. The other direction of Theorem 5.3 follows directly from the Chomsky-Schützenberger theorem, Theorem 5.2, and the closure of C-finite integer sequences under difference.

In Subsection 11.2 we show that the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$ are C-finite for fixed k.

11.1. Proof of Theorem 11.1. Let $a(n)$ satisfy the following recurrence

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

for $n \geq q$ with $p_i \in \mathbb{Z}$. Unwrapping the recurrence we get an expression for $a(n)$ as a sum over all monomials of the form

$$p_{r_1} p_{r_2} \cdots p_{r_k} \cdot a(n - r_1 - \cdots - r_k)$$

where $r_1, \ldots, r_k \in [q]$ and $n - r_1 - \ldots - r_k \in [q]$. We would like to write $a(n)$ as the sum over words of a regular language. Let $\Sigma_1 = [q]$, $\Sigma_2 = \{\tilde{1}, \ldots, \tilde{q}\}$ and $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \{b\}$. Let $L_{rec(q)}$ be the regular language over Σ given by the regular expression

$$(\tilde{1} + b\tilde{2} + bb\tilde{3} + \ldots + b^{q-1}\tilde{q}) \cdot (1 + b2 + bb3 + \ldots + b^{q-1}q)^{\star}.$$

The following function f is a one-to-one map between:

(i) tuples $\bar{r} = (r_1, \ldots, r_k)$ satisfying $n - r_1 - \ldots - r_k \in [q]$ and $r_i \in [q]$ for $i = 1, \ldots, k$.
(ii) words w in $L_{rec(q)}$.

f is given by

$$f(\bar{r}) = b^{t_{\bar{r}}-1}\tilde{t}_{\bar{r}} \cdot (b^{r_1-1}r_1 \cdots b^{r_k-1}r_k),$$

where $t_{\bar{r}} = n - r_1 - \cdots - r_k$. It is not difficult to see that $f(\bar{r})$ belongs to $L_{rec(q)}$ by construction and that f is indeed a bijection between (i) and (ii). The sequence $a(n)$ can now be written explicitly as

$$a(n) \quad = \quad \sum_{w \in L_{rec(q)}} p_1^{|\{j:w[j]=1\}|} \cdots p_q^{|\{j:w[j]=p_q\}|} \cdot a(1)^{|\{j:w[j]=\tilde{1}\}|} \cdots a(q)^{|\{j:w[j]=\tilde{q}\}|} \tag{5}$$

Let $\Sigma_{(1)} = \Sigma - \{1\} \cup \{1_{(1)}, \ldots, 1_{(|p_1|)}\}$ be the alphabet obtained by replacing the letter 1 by $|p_1|$ many new letters. Let $h_1 : \Sigma_{(1)} \to \Sigma^\star$ be given by $h_1(1_{(i)}) = 1$ for all i and $h_1(j) = j$ otherwise. The function h_1 is a homomorphism from $\Sigma_{(1)}$ to $\Sigma^\star$. By the closure of regular languages under inverse homomorphisms, cf. [**44**, Page 61], the language $h^{-1}(L_{rec(q)})$ is regular, where

$$h^{-1}(L_{rec(q)}) = \{x \in \Sigma^\star : h(x) \in L_{rec(q)}\} .$$

It holds that we can replace p_1 in Equation (5) by 1 if we replace $L_{rec(q)}$ by $L_{(1)}$:

$$a(n) \quad = \quad \sum_{w \in L_{(1)}} (sign(p_1))^{|\{j:h(w[j])=1\}|} \cdot (p_2)^{|\{j:w[j]=2\}|} \cdots p_q^{|\{j:w[j]=p_q\}|} \cdot a(1)^{|\{j:w[j]=\tilde{1}\}|} \cdots a(q)^{|\{j:w[j]=\tilde{q}\}|}$$

where $sign(p_1) = 1$ if $p_1 > 0$ and otherwise $sign(p_1) = -1$. Continuing similarly we may replace each $p_1, \ldots, p_q$ and $a(1), \ldots, a(q)$ with 1 or -1 and obtain a regular language L' over

$$\Sigma' = \{j_{(i)} : j \in [q], i \in [|p_j|]\} \cup \{\tilde{j_{(i)}} : j \in [q], i \in [|a(j)|]\} \cup \{b\} ,$$

a homomorphism $\Sigma' \to \Sigma^\star$, $h(j_{(i)}) = j$, and two sets $S_1 \subseteq \Sigma_1$ and $S_2 \subseteq \Sigma_2$ for which

$$a(n) = \sum_{w \in L'} (-1)^{|\{j:h'(w[j]) \in S_1 \cup S_2\}|}$$

For any set of letters $D \subseteq \Sigma'$, the languages $L_{even(D)}$ and $L_{odd(D)}$ over Σ', which consist of words with an even (respectively odd) number of occurrences of the letters of D, are regular. Hence by the closure of regular languages under intersection and union, $a(n)$ can be written in the form $a(n) = d_{L_A}(n) - d_{L_B}(n)$, where $d_{L_A}(n)$ and $d_{L_B}(n)$ are the counting functions of regular languages L_A and L_B.

Finally, let L_B^c be the language obtained by replacing all the letters in the complement of L_B, $\Sigma' - L_B$, by new letters which do not appear in Σ'. Then $d_{L_B}(n) = |\Sigma'|^n - d_{L_B^c}(n)$. Since the alphabets of L_A and L_B^c are disjoint, $d_{L_B^c \cup L_A}(n) = d_{L_B^c}(n) + d_{L_A}(n)$. Therefore, $d_{L_B^c \cup L_A}(n) - |\Sigma'|^n = d_{L_A}(n) - d_{L_B}(n) = a(n)$. By the closure of regular languages under complement and union, $L_B^c \cup L_A$ is a regular language, implying the required result.

. □

11.2. The Stirling Numbers of the Second Kind. The Stirling numbers of the second kind were discussed in Subsection 7.7. Recall the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$ count partitions of $[n]$ into k non-empty parts.

Proposition 11.1. *Let k be fixed. There exists a regular lanuage L_k such that* $\left\{ {n \atop k} \right\} = d_{L_k}(n)$.

PROOF. Let $\Sigma = [k]$ and let L_k be the language which consists of all words over Σ in which every letter of Σ occurs at least once, and furthermore, a letter $i \in [k]$ may only occur in a word w of L_k if $i-1$ occurs before it in w. The language L_k is given by the regular expression

$$1 \cdot 1^\star \cdot 2 \cdot \{1,2\}^\star \cdot 3 \cdot \{1,2,3\}^\star \cdots k \cdot [k]^\star ,$$

where $\{1,2,\dots,i\}^\star$ denotes the set of words with letters $\{1,2,\dots,i\}$.

Let P be the set of all partitions of $[n]$ with exactly k non-empty parts. Let $f : L_k \to P$ be the function given by $f(w) = p_w$, where p_w is the following partition:

$$p_w = \{\{i : w[i] = 1\}, \dots, \{i : w[i] = k\}\} .$$

By the definition of L_k, p_w is indeed a partition of $[n]$ which consists of k non-empty parts. Moreover, f is a bijection, implying $\left\{ {n \atop k} \right\} = d_{L_k}(n)$. $\square$

12. Holonomic Sequences

In this section we discuss an interpretation of SP-recursive and P-recursive integer sequences. We give two logical interpretations and characterizations, one with lattice paths, and one with positionally weighted words. The lattice path approach is more suitable for SP-recursive sequences, whereas in the case of P-recursive sequences which are not SP-recursive weights are more appropriate. We omit the proofs and emphasize the concepts. Missing proofs may be found in [**47, 46**].

12.1. SP-recursive Sequences and Lattice Paths. Various P-recursive sequences have interpretations as counting lattice paths. A prominent example is given by the Catalan numbers $C(n)$, cf. Section 7.8. The Catalan numbers $C(n)$ count the number of paths in an $n \times n$ grid from the lower left corner to the upper right corner which have steps $\to$ and $\uparrow$ and which never go above the diagonal line, see Figure 1. The central binomial coefficient, discussed in Section 7.4, counts paths

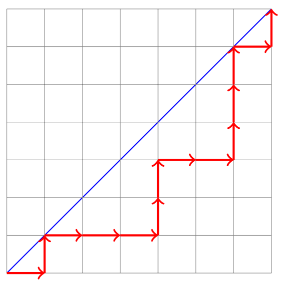

FIGURE 1. A legal path counted by $C(n)$

similar to those counted by $C(n)$, with the exception that the paths are allowed to go above the diagonal. Many other P-recursive sequences which can be interpreted as counting lattice paths can be found in [**69**]. Among them we find the Motzkin numbers and the Schröder numbers.

Possibly the simplest SP-recursive sequence is the factorial $n!$, satisfying the recurrence $(n+1)! = (n+1) \cdot n!$. The factorial was discussed in Section 7.1. $n!$ can be interpreted as counting lattice paths in an $(n+1) \times (n+1)$ grid which:

(Req. a) start from the lower left corner and end at the upper right corner,
(Req. b) consist of steps from $\to, \uparrow, \downarrow$,
(Req. c) do not cross the diagonal line, and
(Req. d) are self-avoiding.

For an example of a path satisfying these requirements see Figure 2. Notice that re-

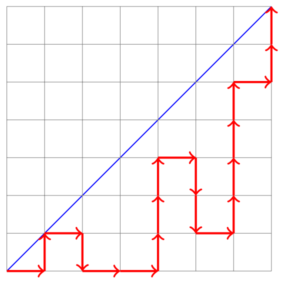

FIGURE 2. A legal path counted by $n!$

moving the requirement that the lattice paths are self-avoiding will mean that there are always infinitely many such paths. Furthermore, notice that this requirement is not needed in the case of the Catalan numbers since the paths are monotonic.

12.2. Combining Lattice Paths with Regular Languages. The lattice paths of interest to us are lattice paths in an $(n+1) \times (n+1)$ grid which satisfy requirements (Req. a), (Req. b), (Req. c) and (Req. d) and an additional requirement (Req. e). We define these lattice paths now.

DEFINITION 12.1. *Let w be a word of length n over an alphabet Σ and let $\sigma \in \Sigma$. A (w, σ)-*path *is a lattice path in an $(n+1) \times (n+1)$ grid which:*

(Req. a) starts from the lower left corner and ends at the upper right corner,
(Req. b) consists of steps from $\to, \uparrow, \downarrow$,
(Req. c) does not cross the diagonal line,
(Req. d) is self-avoiding, and
(Req. e) for any $j \in [n]$, if the j-th letter of w is not σ, then any step starting at (i, j) for any $i \in [n]$ must be a right step $\to$.

We think of the word w as labeling the columns of the grid. In columns labeled with a letter which is not σ, the only step allowed is $\to$. A legal $(ccbcbbb, b)$-path is shown in Figure 3.

DEFINITION 12.2. *Let L be a regular language over Σ and let $\bar{\sigma} = (\sigma_1, \ldots, \sigma_r)$ be a tuple of Σ letters. We define the function $m_{L,\bar{\sigma}} : \mathbb{N} \to \mathbb{N}$ as follows: $m_{L,\bar{\sigma}}(n)$ is the number of tuples $(w, p_1, \ldots, p_r)$ where $w \in L$ and each p_i is a (w, σ_i)-path.*

We say a sequence $a(n)$ has an LP-interpretation *if there exists a regular language L and a tuple of letters $\bar{\sigma}$ such that $a(n) = m_{L,\bar{\sigma}}(n)$.*

THEOREM 12.1 ([**46**]). *Let $a(n)$ be a sequence of integers.*

(i) $a(n)$ is SP-recursive iff $a(n)$ is the difference of two sequences $d_1(n)$ and $d_2(n)$ which have LP-interpretations,

$$a(n) = d_1(n) - d_2(n)\,.$$

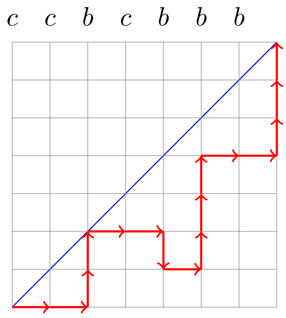

FIGURE 3. A legal $(ccbcbbb, b)$-path

(ii) *$a(n)$ is P-recursive with leading polynomial $p_q(x)$ iff $a(n)$ there exist $d_1(n)$ and $d_2(n)$ which have LP-interpretations such that*

$$a(n) = \frac{d_1(n) - d_2(n)}{\prod_{s=1}^{n} p_q(s)} .$$

12.3. Permutations and Lattice Paths. We now discuss two examples of interpreting SP-recursive sequences which arise from counting two types of permutations as LP-interpretations. In both cases we already know that they are SP-recursive. The point here is to exhibit explicitely how they can be seen as LP-sequences.

12.3.1. *Counting permutations with a fixed number of cycles.* The Stirling numbers of the first kind $\left[{n \atop k}\right]$ were discussed in Section 7.6. They count the number of permutations of $[n]$ with exactly k cycles and are SP-recursive. They satisfy

$$\left[{n+1 \atop k}\right] = n \cdot \left[{n \atop k}\right] + \left[{n \atop k-1}\right] .$$

$\left[{n+1 \atop k}\right]$ can be interpreted naturally as counting (w, σ)-paths.

Let L_k be the set of words over alphabet $\{0, 1\}$ in which 1 occurs exactly $k-1$ times. This is a regular language, given by the regular expression $(0^\star 1)^{k-1} 0^\star$. We will see that

$$m_{L_k,0}(n) = \left[{n+1 \atop k}\right] . \tag{6}$$

By definition, $m_{L_k,0}(n)$ counts (w, σ)-paths, where $w \in L_k$. Let $u \in L_k$ and let $A_u = \{j+1 \mid u[j] = 1\}$. The number of (u, σ)-paths equals

$$\prod_{j:j+1 \notin A_u} j . \tag{7}$$

On the other hand, we want to count the permutations of $[n+1]$ such that $i \in [n+1]$ is the minimal element in its cycle iff $i \in A_u \cup \{1\}$. Let $i \in [n+1]$. Assume we have a permutation π_i of $[i]$ such that the set of elements which are minimal in their cycle in π_i is $(A_u \cup \{1\}) \cap [i]$. We want to count the number of ways of adding the element $i+1$ to π_i and getting a permutation π_{i+1} of $[i+1]$ for which the set of elements which are minimal in their cycle is $(A_u \cup \{1\}) \cap [i+1]$.

If $i+1 \in A_u$ then $i+1$ must be the minimal element in its cycle. This means that $i+1$ must form a new cycle of its own and so there is exactly one permutation π_{i+1} which extends π_i in this way. Otherwise, if $i+1 \notin A_u$ then $i+1$ must not be the minimal element in its cycle in π_{i+1}. Hence, it must be added to one of the existing cycles. There are i ways to do so, which correspond to choosing the element $j \in [i]$ which $i+1$ will follow in π_{i+1}.

We get that the number of permutations of $[n+1]$ for which the set of elements which are minimal in their cycle is $A_u \cup \{1\}$ is given in Equation (7) and is equal to the number of (u,σ)-paths. Summing over words $u \in L_k$ or equivalently, over sets $A = A_u \cup \{1\}$ of size k, we get Equation (6).

12.3.2. *Permutations without fixed points.* The derangement numbers $D(n)$ were defined in Section 7.3. $D(n)$ counts permutations of $[n]$ with no fixed-point. They satisfy the SP-recurrence

$$D(n+1) = n \cdot D(n) + n \cdot D(n-1) \tag{8}$$

with initial conditions $D(0) = 1$ and $D(1) = 0$. Let $D'(n)$ be the sequence such that $D(n+1) = D'(n)$, i.e. $D'(n)$ is the number of permutations of $[n+1]$ without fixed-points. Then

$$D'(n) = n \cdot D'(n-1) + n \cdot D'(n-2) \tag{9}$$

with initial conditions $D'(1) = D(0) = 1$ and $D'(2) = D(1) = 0$.

It can be shown by induction that

$$D'(n) = \sum_{w \in \{a,b,c,d\}^n \cap L_{der}} \prod_{w[i] \in \{a,b\}} i \tag{10}$$

where the summation is over words of length n in L_{der}. The language L_{der} consists of all words w such that $w = d$ or

(i) $w[i] = c$ iff $w[i+1] = b$, and
(ii) $w[1] \cdot w[2] \cdot w[3] = dcb$.

We can think of Equation (10) as the sum over all paths in the recurrence tree of equation (9) from the root to a leaf $D'(1)$ (notice a path in the recurrence tree that ends in $D'(2) = 0$ has value 0). Such a path can be described by $1 = t_1 \leq \ldots \leq t_r = n$ such that for each i, $1 < i \leq r$, the difference of subsequent elements $t_i - t_{i-1}$ is either 1 or 2. The elements t_i in $[n]$ for which a recurrence step of the form $i \cdot D'(i-1)$ was chosen (i.e., those for which $t_i - t_{i-1} = 1$) are assigned letter a, whereas b is assigned to those elements t_i of $[n]$ which correspond to a choice of the form $i \cdot D'(i-2)$ (i.e., those t_i for which $t_i - t_{i-1} = 2$). We assign c to all the elements $i \in [n] - \{t_1, \ldots, t_r\}$, which are skipped by a recursive choice $j \cdot D'(j-2)$, where $j = i+1$. The letter d is assigned to the leafs $D'(1)$. Condition (i) requires that $i-1$ is skipped iff $i \cdot D'(i-2)$ is chosen for i. Condition (ii) requires that the path in the recurrence tree does not end in $D'(2) = 0$, but rather skips from $D'(3)$ to $D'(1)$. Notice this is a regular language, given by the regular expression

$$dcb \cdot \left(a^\star \, (cb)^\star\right)^\star + d$$

Equation (10) can be interpreted as counting the number of tuples (w, p_0, p_1) where $w \in L_{der}$ is of length $|w| = n$, p_0 is a (w,a)-path and p_1 is a (w,b)-path.

12.4. P-recursive sequences and positional weights. In this subsection we present a logical interpretation for P-recursive sequences which uses positional weights. The interpretation sheds light on P-recurrences as a generalization of hypergeometric recurrences.

Recall a sequence of integers $a(n)$ is hypergeometric if for all $n > 1$,

$$a(n) = \frac{p_0(n)}{p_1(n)} a(n-1) \tag{11}$$

where $p_1(x), p_0(x) \in \mathbb{Z}[x]$ are polynomials and $p_1(n)$ does not vanish for any n. A prominent example of a hypergeometric sequence is the binomial coefficient $\binom{n}{k}$ fixed k and $n > k$, given by the recurrence

$$\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k}.$$

In Sections 7.8 and 7.4 we saw that the Catalan numbers $C(n)$ and the central binomial coefficient satisfy hypergeometric recurrences.

Given $a(n)$ which satisfies Equation (11), one can write $a(n)$ explicitly as

$$a(n) = a(1) \cdot \prod_{j=2}^{n} \frac{p_0(j)}{p_1(j)}. \tag{12}$$

We may rewrite Equation (12) as follows:

$$a(n) = \sum_{w \in L_{ba^\star} \cap \{a,b\}^n} \prod_{j:w[j]=b} a(1) \prod_{j:w[j]=a} \frac{p_0(j)}{p_1(j)},$$

where $L_{ba^\star}$ is the language specified by the regular expression $ba^\star$, where $w[j]$ is the j-th letter of w, and where the products range over elements $j \in [n]$ such that $w[j] = b$ or $w[j] = a$ respectively. We will show that any P-recursive sequence can be interpreted in a similar way. Let Σ be an alphabet. For $s \in \Sigma$, let $\alpha_s : \mathbb{N} \to \mathbb{Z}$ be a function. We define the *weight* $\alpha(w)$ *of a word* $w \in \Sigma^\star$ by

$$\alpha(w) = \prod_{j=1}^{|w|} \alpha_{w[j]}(j)$$

where $|w|$ is the length of w. For a language $L \subseteq \Sigma^\star$ we define its *positionally weighted density* by

$$d_{L,\alpha}(n) = \sum_{w \in L \cap \Sigma^n} \alpha(w)$$

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

DEFINITION 12.3. *A sequence $a(n)$ of integers has a* PW-interpretation *if there exists a regular language $L \subseteq \Sigma^\star$, and for each $s \in \Sigma$ a rational function $\alpha_s \in \mathbb{Q}(x)$ such that $a(n) = d_{L,\alpha}(n)$.*

THEOREM 12.2. *Let $a(n)$ be a sequence of integers. Then $a(n)$ is P-recursive iff $a(n)$ has a PW-interpretation.*

Theorem 12.2 can be modified to characterize SP-recursive sequences. In this case one simply needs to restrict the rational functions $\alpha_1, \ldots, \alpha_k$ to be polynomials over the integers.

12.5. Two explicit examples. We now give two examples. The Apéry numbers show how to get a PW-interpretation from its P-recurrence. The derangement numbers illustrate how to use an explicit summation formula for a sequence to get a P-recurrence.

12.5.1. *A Non-trivial Example of a Holonomic Sequence.* The Apéry numbers appear in Apéry's proof that $\zeta(3)$ is irrational and are known to be P-recursive, cf. [**3, 68**]. They satisfy the P-recurrence

$$n^3 b_n = (34n^3 - 51n^2 + 27n - 5)b_{n-1} - (n-1)^3 b_{n-2}\,.$$

The purpose of this subsection is to show how the polynomials of the P-recurrence of $a(n)$ are used to compute the weights in for the PW-interpretation of $a(n)$.

Using a similar argument to the one used in Subsection 12.3.2 it holds that:

$$b_n = \sum_{w \in L_{rec(2)}, |w|=n}$$

$$\left(\prod_{j:w[j]=1} \frac{(34j^3 - 51j^2 + 27j - 5)}{j^3} \cdot \prod_{j:w[j]=2} \frac{-(j-1)^3}{j^3} \prod_{j:w[j]=\tilde{1}} b(1) \prod_{j:w[j]=\tilde{2}} b(2) \right).$$

where the language $L_{rec(2)}$ is from Subsection 11.1 with $q = 2$.

Let

$$\alpha_1(x) = \frac{(34x^3 - 51x^2 + 27x - 5)}{x^3}, \quad \alpha_2(x) = \frac{-(x-1)^3}{x^3},$$

$\alpha_{\tilde{1}}(x) = b_1$, $\alpha_{\tilde{2}}(x) = b_2$, and $\alpha_b(x) = 1$. b_n has the following PW-interpretation:

$$b_n = \sum_{w \in L_{rec(2)}, |w|=n} \left(\prod_{j=1}^{|w|} \alpha_{w[j]}(j) \right) = d_{L_{rec(2)}}(n), \alpha\,.$$

12.5.2. *Derangement numbers again.* The derangement numbers are given explicitly by the formula

$$D(n) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}\,.$$

Using this formula one can easily see that $D(n)$ has a PW-interpretation,

$$D(n) = \sum_{w \in L_{0^\star 1^\star}, |w|=n} \left(\prod_{j:w[j]=0} -1 \prod_{j:w[j]=1} j \right),$$

where $L_{0^\star 1^\star}$ is the language obtained from the regular expression $0^\star 1^\star$.

Part 4. Modular Recurrence relations

13. *DU*-index and Specker index

In this section we give the detailed definitions of the finiteness conditions mentioned in Section 4. Specker's proof in [**66**] of Theorem 5.8 is based on the analysis of an equivalence relation $\sim_{\mathcal{C}}$ induced by a class of structures $\mathcal{C}$. However, we first look at a simpler case of disjoint unions of structures.

13.1. DU-index of a class of structures.

We denote by $\mathfrak{A} \sqcup \mathfrak{B}$ the disjoint union of two $\bar{R}$-structures $\mathfrak{A}$ and $\mathfrak{B}$.

Definition 13.1. *Let $\mathcal{C}$ be a class of $\bar{R}$-structures.*

(i) We say that $\mathfrak{A}_1$ is $DU(\mathcal{C})$-equivalent to $\mathfrak{A}_2$ and write $\mathfrak{A}_1 \sim_{DU(\mathcal{C})} \mathfrak{A}_2$, if for every $\bar{R}$-structure $\mathfrak{B}$, $\mathfrak{A}_1 \sqcup \mathfrak{B} \in \mathcal{C}$ if and only if $\mathfrak{A}_2 \sqcup \mathfrak{B} \in \mathcal{C}$.

(ii) The DU-index of $\mathcal{C}$ is the number of $DU(\mathcal{C})$-equivalence classes.

DEFINITION 13.2. *A class of structures $\mathcal{C}$ is a* Gessel class *if for every $\mathfrak{A}$ and $\mathfrak{B}$, $\mathfrak{A} \sqcup \mathfrak{B} \in \mathcal{C}$ iff both $\mathfrak{A} \in \mathcal{C}$ and $\mathfrak{B} \in \mathcal{C}$.*

I. Gessel in [**38**, Theorem 4.2] looks at Gessel classes of directed graphs which in addition have a bounded degree. He proves the following congruence theorem:

THEOREM 13.1 (I. Gessel 1984). *If $\mathcal{C}$ is a Gessel class of directed graphs of degree at most d, then*

$$d_{\mathcal{C}}(m+n) \equiv d_{\mathcal{C}}(m) \cdot d_{\mathcal{C}}(n) \pmod{\frac{m}{\ell}}$$

where ℓ is the least common multiple of all divisors of m not greater than d.

In particular, $d_{\mathcal{C}}(n)$ satisfies for every $m \in \mathbb{N}$ the linear recurrence relation $d_{\mathcal{C}}(n) \equiv a^{(m)} d_{\mathcal{C}}(n - d!m) \pmod{m}$ where $a^{(m)} = d_{\mathcal{C}}(d!m)$.

A less informative version of this theorem was stated in the introduction as Theorem 5.9.

In its proof, the following simple observation was implicitly used:

OBSERVATION 13.2. *If $\mathcal{C}$ is a Gessel class of $\bar{R}$-structures then $\mathcal{C}$ has DU-index at most 2.*

PROOF. We observe that all members of $\mathcal{C}$ are in one equivalence class, as well as that all other $\bar{R}$-structures are in one equivalence class (which is usually different than that of the members of $\mathcal{C}$). $\square$

REMARK 13.1. *The converse is not true, because if $\mathcal{C}$ is a class of connected graphs then $\mathcal{C}$ has DU-index 1, but is not Gessel.*

This may seem unnatural, as we would expect to have at least two classes, the connected graphs and the non-connected graphs. If we allowed the empty structure to be a structure, this would indeed be the case. But in logic and model theory empty structures are traditionally avoided, because $\forall x \phi(x) \to \exists x \phi(x)$ is considered a tautology. This is in contrast to graph theory, where the empty graph is allowed.

Theorem 5.10 of the introduction can be viewed as a strong variation of Gessel's Theorem for arbitrary $\bar{R}$-structures of bounded degree. Note that the formulation of Gessel's Theorem as Theorem 13.1 contains much more information on the recurrence relation than Theorem 5.10.

13.2. Substitution of structures. A pointed $\bar{R}$-structure is a pair $(\mathfrak{A}, a)$, with $\mathfrak{A}$ an $\bar{R}$-structure and a an element of the universe A of $\mathfrak{A}$. In $(\mathfrak{A}, a)$, we speak of the structure $\mathfrak{A}$ and the *context* a.

The terminology is borrowed from the terminology used in dealing with tree automata, cf. [**63, 35**].

DEFINITION 13.3. *Given two pointed structures $(\mathfrak{A}, a)$ and $(\mathfrak{B}, b)$ we form a new pointed structure $(\mathfrak{C}, c) = Subst((\mathfrak{A}, a), (\mathfrak{B}, b))$ defined as follows:*

(i) The universe of $\mathfrak{C}$ is $A \cup B - \{a\}$.
(ii) The context c is given by b, i.e., $c = b$.
(iii) For $R \in \bar{R}$ of arity r R^C is defined by

$$R^C = (R^A \cap (A - \{a\})^r) \cup R^B \cup I$$

where for every relation in R^A which contains a, I contains all possibilities for replacing these occurrences of a with (identical or differing) members of B.

(iv) We similarly define $Subst((\mathfrak{A}, a), \mathfrak{B})$ for a structure $\mathfrak{B}$ that is not pointed, in which case the resulting structure $\mathfrak{C}$ is also not pointed.

Definition 13.4. *Let $\mathcal{C}$ be a class of $\bar{R}$-structures.*

(i) We define an equivalence relation between $\bar{R}$-structures; we say that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are equivalent, denoted $\mathfrak{B}_1 \sim_{Su(\mathcal{C})} \mathfrak{B}_2$, if for every pointed structure $(\mathfrak{A}, a)$ we have that $Subst((\mathfrak{A}, a), \mathfrak{B}_1) \in \mathcal{C}$ if and only if $Subst((\mathfrak{A}, a), \mathfrak{B}_2) \in \mathcal{C}$.

(ii) The Specker index *of $\mathcal{C}$ is the number of equivalence classes of $\sim_{Su(\mathcal{C})}$.*

The Specker index is related to the DU-index by the following.

Proposition 13.3. *Let $\mathcal{C}$ be a class of $\bar{R}$-structures and $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two $\bar{R}$-structures.*

(i) If $\mathfrak{A}_1 \sim_{Su(\mathcal{C})} \mathfrak{A}_2$, then $\mathfrak{A}_1 \sim_{DU(\mathcal{C})} \mathfrak{A}_2$.

(ii) The Specker index of $\mathcal{C}$ is at least as big as the DU-index of $\mathcal{C}$. In particular, if the Specker index of $\mathcal{C}$ is finite, then so is its DU-index.

We also have in analogy to Observation 13.2,

Observation 13.4. *If $\mathcal{C}$ is a class of pointed $\bar{R}$-structures such that $Subst((\mathfrak{A}, a), (\mathfrak{B}, b)) \in \mathcal{C}$ iff both $(\mathfrak{A}, a), (\mathfrak{B}, b) \in \mathcal{C}$ then the Specker-index of $\mathcal{C}$ is 2.*

Specker's proof of Theorem 5.8 consist of a purely combinatorial part:

Lemma 13.5 (Specker's Lemma). *Let $\mathcal{C}$ be a class of $\bar{R}$-structures of finite Specker index with all the relation symbols in $\bar{R}$ of arity at most 2. Then $d_{\mathcal{C}}(n)$ satisfies modular linear recurrence relations for every $m \in \mathbb{N}$.*

The proof will be given in section 16.

13.3. Classes of finite Specker or DU-index. Using Proposition 13.3 we have seen that Gessel classes have finite DU-index, and that all classes of connected graphs have finite DU-index. We shall now exhibit a class $\mathcal{C}$ that has DU-index at most 2, but has an infinite Specker index. As stated in Section 7.4, $EQ_2CLIQUE$ denotes the class of graphs which consist of two disjoint cliques of equal size.

Proposition 13.6. *The class $EQ_2CLIQUE$ has infinite Specker index.*

Proof. We show that for all $i, j \in \mathbb{N}, 1 \leq i < j$ the pairs of cliques K_i and K_j are inequivalent with respect to $\sim_{Su(EQ_2CLIQUE)}$. To see this we look at the pointed structure $(K_j \sqcup K_1, a)$ where the vertex of K_1 is the distinguished point a. Substituting K_j for a gives a disjoint union of two K_j's, whereas substituting K_i, $i \neq j$ for a gives a disjoint union of $K_j \sqcup K_i$. The former is in $EQ_2CLIQUE$ whereas the latter is not, which proves our claim. □

Let the class $\overline{EQ_2CLIQUE}$ be the class of all the complement graphs of members of $EQ_2CLIQUE$. We note that $\overline{EQ_2CLIQUE}$ contains graphs of arbitrary large degree which are all connected.

Corollary 13.7. *The class $\overline{EQ_2CLIQUE}$ has finite DU-index, but infinite Specker index.*

Proof. As graphs in $\overline{EQ_2CLIQUE}$ are all connected, $\overline{EQ_2CLIQUE}$ has DU-index at most 2, using Proposition 13.1. On the other hand, it is not hard to see that similarly to $EQ_2CLIQUE$ this class has an infinite Specker index. $\Box$

It is an easy exercise to show that the class HAM of graphs which contain a Hamiltonian cycle also has infinite Specker index. We have seen in Subsection 3.2 that the classes HAM, $EQ_2CLIQUE$ and $\overline{EQ_2CLIQUE}$ are not **CMSOL**-definable. So far, our examples of the classes of infinite Specker index were not definable in **CMSOL**. This is no accident. Specker noted that all **MSOL**-definable classes of $\bar{R}$-structures have a finite Specker index. We shall see that this can be improved.

Theorem 13.8. *If $\mathcal{C}$ is a class of $\bar{R}$-structures (with no restrictions on the arity) which is* **CMSOL***-definable, then $\mathcal{C}$ has a finite Specker index.*

The proof is given in Section 14. It uses a form of the Feferman-Vaught Theorem for **CMSOL**.

13.4. A continuum of classes of finite Specker index. As there are only countably many regular languages over a fixed alphabet, the Myhill-Nerode theorem implies that there are only countably many languages with finite OS-index. In contrast to this, for general relational structures, there are plenty of classes of graphs which have finite Specker index.

Definition 13.5. *Let C_n denote the cycle of size n, i.e. a regular connected graph of degree 2 with n-vertices. Let $A \subseteq \mathbb{N}$ be any set of natural numbers and $Cycle(A) = \{C_n : n \in A\}$.*

Proposition 13.9 (Specker). *$Cycle(A)$ has Specker index at most 5.*

Proof. All binary structures with three or more vertices fall into two classes, the class of graphs G for which $Subst((\mathfrak{A}, a), G) \in Cycle(A)$ if and only if $\mathfrak{A}$ has a single element a (this equals the class $Cycle(A)$), and the class of graphs G for which $Subst((\mathfrak{A}, a), G) \in Cycle(A)$ never occurs (which contains all binary structures which are not graphs, and all graphs with at least three elements which are not in $Cycle(A)$). Binary structures with less than three vertices which are not graphs also fall into the second class above, while the three possible graphs with less then three vertices may form classes by themselves (depending on A). $\Box$

Corollary 13.10 (Specker). *There is a continuum of classes (of graphs, of $\bar{R}$-structures) of finite Specker index which are not* **CMSOL***-definable.*

Proof. Clearly there is continuum of classes of the type $Cycle(A)$, and hence a continuum of classes that are not definable in **CMSOL** (or even in second order logic, **SOL**). $\Box$

It is easy to compute $d_{Cycle(A)}$:

$$d_{Cycle(A)}(n) = \begin{cases} 0 & \text{if } n \notin A \\ (n-1)! & \text{otherwise .} \end{cases}$$

Hence it is trivially MC-finite. This does not have to be necessarily the case. Here is a way to modify the above example.

THEOREM 13.11. *Let $\mathcal{C}$ be a class of $\bar{R}$ structures with counting function $d_{\mathcal{C}}(n)$. For every $A \subseteq \mathbb{N}$ there is a class of structures $\mathcal{C}(A)$ (with all classes being different for different choices of A) such that for all $n \in \mathbb{N}$*

$$d_{\mathcal{C}(A)}(n) = d_{\mathcal{C}}(n) + d_{Cycle(A)}(n)$$

Furthermore, if $\mathcal{C}$ is of finite Specker index, then so is $\mathcal{C}(A)$ for every A (but its index is possibly bigger).

PROOF. The structures in $\mathcal{C}(A)$ have besides the relation symbols $\bar{R}$ a new unary relation symbol U and a new binary relation symbol E. The interpretations of U^B and E^B in a structure $\mathfrak{B} \in \mathcal{C}(A)$ on the universe $[n]$ are given as follows:

(i) Either $U^B = \emptyset$ and $E^B = \emptyset$ and the underlying $\bar{R}$-structure is in $\mathcal{C}$, or
(ii) $U^B = [n]$ and $\langle [n], E^B \rangle \in Cycle(A)$ and all $S^B = \emptyset$ for all $S \in \bar{R}$.

Clearly, we now have $d_{\mathcal{C}(A)}(n) = d_{\mathcal{C}}(n) + d_{Cycle(A)}(n)$.

To see that the Specker index of $\mathcal{C}(A)$ is finite, we look at two $(\bar{R} \cup \{U, E\})$-structures $\mathfrak{B}_1$ and $\mathfrak{B}_2$ and a pointed $(\bar{R} \cup \{U, E\})$-structures $(\mathfrak{A}, a)$, and put $\mathfrak{C}_i = Subst((\mathfrak{A}, a), \mathfrak{B}_i)$ for $i = 1, 2$. Whether $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are $\mathcal{C}(A)$-equivalent can now be decided by checking whether the interpretations of U, E or $S \in \bar{R}$ are empty or full, and whether the corresponding reducts are $\mathcal{C}$-equivalent or $Cycle(A)$-equivalent. $\square$

REMARK 13.2. *This shows that, in contrast to the Myhill-Nerode Theorem, no characterization of the classes of finite Specker index in terms of their definability in* **CMSOL** *is possible.*

13.5. An SOL but not CMSOL-definable class of finite Specker index. Although a classification of all classes of finite Specker index seems unachievable on account of there being a continuum of these, one could still hope to characterize all **SOL**-definable classes of finite index. We show here that definability in **CMSOL** is not such a characterization.

DEFINITION 13.6. *We look at the infinite graph whose vertex set is $\mathbb{Z} \times \mathbb{Z}$ and for which every (i, j) and (i', j') are connected if and only if $|i-i'|+|j-j'| = 1$. We say that a graph G is a* grid graph *if it is a (finite) subgraph of the above infinite graph.*

In [**62**] the following was proven

THEOREM 13.12 ([**62**]). *The class of all grid graphs, which is definable in* **SOL**, *is not definable in* **CMSOL**.

On the other hand, this is also a class with a finite Specker index.

PROPOSITION 13.13. *The class of all grid graphs has a finite Specker index; therefore, there exist classes of finite Specker index which are definable in* **SOL** *but not in* **CMSOL**.

PROOF. We observe that all graphs with five or more vertices fall into the following two Specker equivalence classes:

(i) Graphs G for which $Subst((S, s), G)$ is a grid graph if and only if S is a grid graph and s is an isolated vertex of S.
(ii) Graphs G for which $Subst((S, s), G)$ is never a grid graph.

All binary structures which are not graphs clearly fall into the second equivalence class above. Thus the index is finite. $\square$

14. The rôle of logic

Although Theorem 5.8 is stated for classes of structures definable in some logic, logic is only used to verify the hypothesis of Specker's Lemma, 13.5. In this section we develop the machinery which serves this purpose. The crucial property needed to prove Theorem 13.8 is a reduction property which says that both for the disjoint union $\mathfrak{A} \sqcup \mathfrak{B}$ and for the substitution $Subst((\mathfrak{A}, a), \mathfrak{B})$ the truth value of a sentence $\phi \in \mathbf{CMSOL}(\bar{R})$ depends only on the truth values of the sentences of the same quantifier rank in the structures $\mathfrak{A}$ and $\mathfrak{B}$, respectively $\langle \mathfrak{A}, a \rangle$ and $\mathfrak{B}$. For the case of **MSOL** this follows either from the Feferman-Vaught Theorem for disjoint unions together with some reduction techniques, or using Ehrenfeucht-Fraïssé games. The latter is used in [**66**]. We shall use the former, as it is easier to adapt for **CMSOL**. For the Feferman-Vaught Theorem the reader is referred to [**30**], or the monographs [**55, 43**], or the survey [**51**].

14.1. Quantifier rank. We define the quantifier rank $qr(\phi)$ of a formula ϕ of $\mathbf{CMSOL}(\bar{R})$ inductively as usual: For quantifier free formulas ϕ we have $qr(\phi) = 0$. For boolean operations we take the maximum of the quantifier ranks. Finally, $qr(\exists U \phi) = qr(\exists x \phi) = qr(C_{p,q} x \phi) = qr(\phi) + 1$. We denote by $\mathbf{CMSOL}^q(\bar{R}, \bar{x}, \bar{U})$ the set of $\mathbf{CMSOL}(\bar{R})$-formulas with free variables $\bar{x}$ and $\bar{U}$ which are of quantifier rank at most q. When there are no free variables we write $\mathbf{CMSOL}^q(\bar{R})$.

We write $\mathfrak{A} \sim^q_{\mathbf{CMSOL}} \mathfrak{B}$ for two $\bar{R}$-structures $\mathfrak{A}$ and $\mathfrak{B}$ if they satisfy the same $\mathbf{CMSOL}^q(\bar{R})$-sentences.

The following is folklore, cf. [**25**].

PROPOSITION 14.1. *There are, up to logical equivalence, only finitely many formulas in* $\mathbf{CMSOL}^q(\bar{R}, \bar{x}, \bar{U})$. *In particular, the equivalence relation* $\sim^q_{\mathbf{CMSOL}}$ *is of finite index.*

14.2. A Feferman-Vaught Theorem for CMSOL. We are now interested in how the truth of a sentence in **CMSOL** in the disjoint union of two structures $\mathfrak{A} \sqcup \mathfrak{B}$ depends on the truth of other properties expressible in **CMSOL** which hold in $\mathfrak{A}$ and $\mathfrak{B}$ separately.

The following was first proven by E. Beth in 1952 and then generalized by Feferman and Vaught in 1959 for **FOL**. For **MSOL** it is due to Läuchli and Leonhard, 1966 and for **CMSOL** it is due to B. Courcelle, 1990. The respective references are [**10, 29, 30, 48, 20**].

THEOREM 14.2 (Feferman-Vaught-Courcelle).

(i) For every formula $\phi \in \mathbf{CMSOL}^q(\tau)$ *one can compute in* polynomial time *a sequence of formulas*

$$\langle \psi_1^A, \ldots, \psi_m^A, \psi_1^B, \ldots, \psi_m^B \rangle \in \mathbf{CMSOL}^q(\tau)^{2m}$$

and a boolean function $B_\phi : \{0,1\}^{2m} \to \{0,1\}$ *such that*

$$\mathfrak{A} \sqcup \mathfrak{B} \models \phi$$

if and only if

$$B_\phi(b_1^A, \ldots b_m^A, b_1^B, \ldots b_m^B) = 1$$

where $b_j^A = 1$ *iff* $\mathfrak{A} \models \psi_j^A$ *and* $b_j^B = 1$ *iff* $\mathfrak{B} \models \psi_j^B$.

A detailed proof is found in [**20**, Lemma 4.5, page 46ff].

14.3. Quantifier-free transductions and CMSOL. Let $\bar{R} = R_1, \ldots, R_s$ where R_i is of arity $\rho(i)$. An $\bar{R}$-translation scheme Φ is a sequence of quantifier-free formulas $\Phi = \langle \theta_0(x), \theta_i(x_1, \ldots, x_{\rho(i)} : i \leq s \rangle$ with free variables as indicated. With Φ we associate a map Φ^* which maps an $\bar{R}$-structure $\mathfrak{A}$ to an $\bar{R}$-structure where the universe is the subset of the universe of $\mathfrak{A}$ defined by θ_0, and where the interpretations of R_i are replaced by the relations defined by θ_i. Φ^* is called a *quantifier free $\bar{R}$-transduction.*

For the general framework of translation schemes and transductions, cf. [**21, 49, 25**]. Note that θ_0 has only one free variable. In the literature this corresponds to *scalar transductions.*

LEMMA 14.3. *Let Φ^* be a quantifier free (scalar) $\bar{R}$-transduction. Assume $\mathfrak{A}_1, \mathfrak{A}_2$ are $\bar{R}$-structures and $\mathfrak{A}_1 \sim^q_{\mathbf{CMSOL}} \mathfrak{A}_2$. Then $\Phi^*(\mathfrak{A}_1) \sim^q_{\mathbf{CMSOL}} \Phi^*(\mathfrak{A}_2)$.*

All we need here is that substitution of pointed structures in pointed structures can be obtained from the disjoint union of the two pointed structures by a scalar transduction. Note that the disjoint union of two pointed structures is, strictly speaking, "doubly pointed", and the two distinguished points play different rôles.

LEMMA 14.4. *$Subst((\mathfrak{A}, a), (\mathfrak{B}, b))$ can be obtained from the doubly pointed disjoint union of $(\mathfrak{A}, a)$ and $(\mathfrak{B}, b))$ by a quantifier free transduction.*

SKETCH OF PROOF: The universe of the structure is $C = (A \sqcup B) - \{a\}$. For each relation symbol $R \in \bar{R}$ we put

$$R^C = R^A|_{A-\{a\}} \cup R^B \cup \{(a', b) : (a', a) \in R^A, b \in B\}$$

This is clearly expressible as a quantifier free transduction from the disjoint union. □

PROPOSITION 14.5. *Assume $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1, \mathfrak{B}_2$ are $\bar{R}$-structures and with contexts a_1, a_2, b_1, b_2, respectively, and*

$$(\mathfrak{A}_1, a_1) \sim^q_{\mathbf{CMSOL}} (\mathfrak{A}_2, a_2) \text{ and } (\mathfrak{B}_1, b_1) \sim^q_{\mathbf{CMSOL}} (\mathfrak{B}_2, b_2).$$

Then $Subst((\mathfrak{A}_1, a_1), (\mathfrak{B}_1, b_1)) \sim^q_{\mathbf{CMSOL}} Subst((\mathfrak{A}_2, a_2), (\mathfrak{B}_2, b_2))$.

PROOF. Use Theorem 14.2, Lemma 14.3 and Lemma 14.4. □

14.4. Finite index theorem for CMSOL. Now we can state and prove the Finite Index Theorem;

THEOREM 14.6. *Let $\mathcal{C}$ be defined by a $\mathbf{CMSOL}(\bar{R})$-sentence ϕ of quantifier rank q. Then $\mathcal{C}$ has finite Specker index, and also finite DU-index, which are both bounded by the number of inequivalent $\mathbf{CMSOL}^q(\bar{R})$-sentences. This number is finite by Proposition 14.1.*

PROOF. We have to show that the equivalence relation $\mathfrak{A} \sim^q_{\mathbf{CMSOL}} \mathfrak{B}$ is a refinement of $\mathfrak{A} \sim_{Su(\mathcal{C})} \mathfrak{B}$. But this follows from Proposition 14.5. For the *DU*-index this follows from Proposition 13.3. □

PROBLEM 2. *Are there any logics $\mathcal{L}$ on finite structures extending* **CMSOL** *such that the analog of Theorem 14.6 remains true?*

15. Structures of bounded degree

DEFINITION 15.1. *For an* **MSOL** *class* $\mathcal{C}$*, denote by*[3] $f_{\mathcal{C}}^{(d)}(n)$ *the number of structures over* $[n]$ *that are in* $\mathcal{C}$ *and whose degree is at most* d.

In this section we prove Theorem 5.10 in the following form:

THEOREM 15.1. *If* $\mathcal{C}$ *is a class of* $\bar{R}$*-structures which has a finite DU-index, then the function* $f_{\mathcal{C}}^{(d)}(n)$ *is ultimately periodic modulo* m*, for every* $m \in \mathbb{N}$*, and therefore is MC-finite.*

Furthermore, if all structures of $\mathcal{C}$ *are connected, then* $f_{\mathcal{C}}^{(d)}(n)$ *is ultimately zero modulo* m*, and therefore is trivially MC-finite.*

LEMMA 15.2. *If* $\mathfrak{A} \sim_{DU(\mathcal{C})} \mathfrak{B}$*, then for every* $\mathfrak{C}$ *we have*

$$\mathfrak{C} \sqcup \mathfrak{A} \sim_{DU(\mathcal{C})} \mathfrak{C} \sqcup \mathfrak{B}.$$

PROOF. Easy, using the associativity of the disjoint union. □

To prove Theorem 15.1 we define orbits for permutation groups rather than for single permutations.

DEFINITION 15.2. *Given a permutation group* G *that acts on* A *(and in the natural manner acts on models over the universe* A*), the* orbit *in* G *of a model* $\mathfrak{A}$ *with the universe* A *is the set* $\mathrm{Orb}_G(\mathfrak{A}) = \{\sigma(\mathfrak{A}) : \sigma \in G\}$.

For $A' \subset A$ we denote by $S_{A'}$ the group of all permutations for which $\sigma(u) = u$ for every $u \notin A'$. The following lemma is useful for showing linear congruences modulo m.

LEMMA 15.3. *Given* $\mathfrak{A}$*, if a vertex* $v \in A - A'$ *has exactly* d *neighbors in* A'*, then* $|\mathrm{Orb}_{S_{A'}}(\mathfrak{A})|$ *is divisible by* $\binom{|A'|}{d}$.

PROOF. Let N be the set of all neighbors of v which are in A', and let $G \subset S_{A'}$ be the subgroup $\{\sigma_1\sigma_2 : \sigma_1 \in S_N \wedge \sigma_2 \in S_{A'-N}\}$; in other words, G is the subgroup of the permutations in $S_{A'}$ that in addition send all members of N to members of N. It is not hard to see that $|\mathrm{Orb}_{S_{A'}}(\mathfrak{A})| = \binom{|A'|}{|N|}|\mathrm{Orb}_G(\mathfrak{A})|$. □

The following simple observation is used to enable us to require in advance that all structure in $\mathcal{C}$ have a degree bounded by d.

OBSERVATION 15.4. *We denote by* $\mathcal{C}_d$ *the class of all members of* $\mathcal{C}$ *that in addition have bounded degree* d*. If* $\mathcal{C}$ *has a finite DU-index then so does* $\mathcal{C}_d$. □

Instead of $\mathcal{C}$ we look at $\mathcal{C}_d$, which by Observation 15.4 also has a finite DU-index. We now note that there is only one equivalence class containing any structures whose maximum degree is larger than d, which is the class $\mathcal{N}_{\mathcal{C}}^{(d)} = \{\mathfrak{A} : \forall_{\mathfrak{B}} (\mathfrak{B} \sqcup \mathfrak{A}) \not\models \mathcal{C}_d)\}$ In order to show that $f_{\mathcal{C}}^{(d)}(n)$ is ultimately periodic modulo m, we show a linear recurrence relation modulo m on the vector function $(d_{\mathcal{E}}(n))_{\mathcal{E}}$ where $\mathcal{E}$ ranges over all other equivalence classes with respect to $\mathcal{C}_d$.

Let $C = md!$. We note that for every $t \in \mathbb{N}$ and $0 < d' \leq d$, m divides $\binom{tC}{d'}$. This with Lemma 15.3 allows us to prove the following.

[3]We change here the notation and use $f^{(d)}$ rather than $d^{(d)}$ to avoid confusion of the counting function with the degree.

LEMMA 15.5. *Let $\mathcal{D} \neq \mathcal{N}_\phi$ be an equivalence class for ϕ, that includes the requirement of the maximum degree not being larger than d. Then*

$$d_{\mathcal{D}}(n) \equiv \sum_{\mathcal{E}} a_{\mathcal{D},\mathcal{E},m,(n \bmod C)} d_{\mathcal{E}}(C\lfloor \frac{n-1}{C} \rfloor) \pmod{m},$$

for some fixed appropriate $a_{\mathcal{D},\mathcal{E},m,(n \bmod C)}$.

PROOF. Let $t = \lfloor \frac{n-1}{C} \rfloor$. We look at the set of structures in $\mathcal{D}$ with the universe $[n]$, and look at their orbits with respect to $S_{[tC]}$. If a model $\mathfrak{A}$ has a vertex $v \in [n] - [tC]$ with neighbors in $[tC]$, let us denote the number of its neighbors by d'. Clearly $0 < d' \leq d$, and by Lemma 15.3 the size of $\mathrm{Orb}_{S_{[tC]}}(\mathfrak{A})$ is divisible by $\binom{tC}{d'}$, and therefore it is divisible by m. Therefore, $d_{\mathcal{D}}(n)$ is equivalent modulo m to the number of structures in $\mathcal{D}$ with the universe $[n]$ that in addition have no vertices in $[n] - [tC]$ with neighbors in $[tC]$.

We now note that any such structure can be uniquely written as $\mathfrak{B} \sqcup \mathfrak{C}$ where $\mathfrak{B}$ is any structure with the universe $[n - tC]$, and $\mathfrak{C}$ is any structure over the universe $[tC]$. We also note using Lemma 15.2 that the question as to whether $\mathfrak{A}$ is in $\mathcal{D}$ depends only on the equivalence class of $\mathfrak{C}$ and on $\mathfrak{B}$ (whose universe size is bounded by the constant C). By summing over all possible $\mathfrak{B}$ we get the required linear recurrence relation (cases where $\mathfrak{C} \in \mathcal{N}_{\mathcal{C}}^{(d)}$ do not enter this sum because that would necessarily imply $\mathfrak{A} \in \mathcal{N}_{\mathcal{C}}^{(d)} \neq \mathcal{D}$). $\Box$

PROOF OF THEOREM 15.1: We use Lemma 15.5: Since there is only a finite number of possible values modulo m to the finite dimensional vector $(d_{\mathcal{E}}(n))_{\mathcal{E}}$, the linear recurrence relation in Lemma 15.5 implies ultimate periodicity for n's which are multiples of C. From this the ultimate periodicity for other values of n follows, since the value of $(d_{\mathcal{E}}(n))_{\mathcal{E}}$ for an n which is not a multiple of C is linearly related modulo m to the value at the nearest multiple of C.

Finally, if all structures are connected we use Lemma 15.3. Given $\mathfrak{A}$, connectedness implies that there exists a vertex $v \in A'$ that has neighbors in $A - A'$. Denoting the number of such neighbors by d_v, we note that $|\mathrm{Orb}_{S'_A}(\mathfrak{A})|$ is divisible by $\binom{|A'|}{d_v}$, and since $1 \leq d_v \leq d$ (using $|A'| = tC$) it is also divisible by m. This makes the total number of models divisible by m (remember that the set of all models with $A = [n]$ is a disjoint union of such orbits), so $f_{\mathcal{C}}^{(d)}(n)$ ultimately vanishes modulo m. $\Box$

16. Structures of unbounded degree

In this section we prove Specker's Lemma 13.5 for structures of unbounded degree. In fact this is a somewhat modified version of Specker's simplified proof for the case where $m = p$ is a prime, as in [**67**].

In order to prove that counting functions of classes of finite Specker index (over unary and binary relation symbols) are ultimately periodic modulo any integer m, it is enough to prove this for any $m = p^k$ where p is a prime number; any other m will then follow by using the Chinese Reminder Theorem. In the following subsections we prove ultimate periodicity for $m = p^k$. First we define a permutation group $G_{p,k}$ which ensures that all structures have large orbits under it, apart from those structures which are "invariant enough" to be represented in terms of a sequence of substitutions in a smaller structure. Then, using this group we show a linear recurrence relation in a vector function that is related to our class $\mathcal{C}$, from which Specker's Lemma follows.

16.1. A permutation group ensuring large orbits. To deal with the exceptional case $p = 2$ we let $\tilde{p} = 4$ if $p = 2$, and $\tilde{p} = p$ otherwise. As our structures have only binary and unary relations, we use the language of graphs, and speak of vertices and edges.

In the following we construct a permutation group $G_{p,k}$.

It acts on the set $\{1, \ldots, \tilde{p}^k\}$ and satisfies the following properties:

(i) The size of $G_{p,k}$ is a power of p and hence the size of an orbit of any structure over $\{1, \ldots, n\}$, where $n \geq \tilde{p}^k$, is also a power of p.
(ii) If the orbit of a binary structure $\mathfrak{A}$ with universe $\{1, \ldots, n\}$ has size less than p^k, then $\mathfrak{A}$ is the result of substituting the $\tilde{p}^{k-1}$ many substructures induced on the sets $\{1 + \tilde{p}i, \ldots, \tilde{p}(i + 1)\}$ (a substructure for every $0 \leq i < \tilde{p}^{k-1}$) into a smaller structure. Equivalently, for any set $v \in \{1 + \tilde{p}i, \ldots, \tilde{p}(i + 1)\}$, the relations between v and the vertices outside $\{1 + \tilde{p}i, \ldots, \tilde{p}(i + 1)\}$ are invariant with regards to permuting this subset.

To achieve this we define $G_{p,k}$ is as follows:

(i) We relabel the vertices $\{1, \ldots, \tilde{p}^k\}$ of $\mathfrak{A}$ with vectors of $(\mathbb{Z}_{\tilde{p}})^k$, by relabeling i with $(x_1, \ldots, x_k)$, where $x_j \equiv \lfloor (i-1)/\tilde{p}^{j-1} \rfloor \pmod{\tilde{p}}$ for $1 \leq j \leq k$.
(ii) We define σ_i by

$$\sigma_i((x_1, \ldots, x_k)) = \begin{cases} (x_1, \ldots, x_{i-1}, x_i + 1, x_{i+1}, \ldots, x_k) & \text{if } x_{i+1} = \cdots = x_k = 0 \\ (x_1, \ldots, x_k) & \text{otherwise} \end{cases}$$

with the addition being modulo $\tilde{p}$.
(iii) $G_{p,k}$ is the group generated by $\sigma_1, \ldots, \sigma_k$.

Observation 16.1. *The following are easy to verify:*

(i) $G_{p,k-1}$ is a subgroup of $G_{p,k}$ in the appropriate sense.
(ii) The order of $G_{p,k}$ (and hence of any orbit it induces on structures) is a power of p (remember that $\tilde{p}$ is a power of p).

Some additional terminology is needed:

(i) The set of all vertices labeled by $(x_1, \ldots, x_k)$ where $x_{i+1} = \cdots = x_k = 0$ is called the *i-origin*.
(ii) The set of all vertices labeled by $(x_1, \ldots, x_k)$ where $x_k = r$ for some fixed r is called the *r-shifted $k-1$-origin*;
(iii) Similarly, shifts of other origins: the $(r_1, \ldots, r_l)$-shifted $k-l$ origin is just the set of vertices labeled by $(x_1, \ldots, x_k)$ where $x_i + l = r_i$ for $1 \leq i \leq l$.
(iv) The (0-shifted) 1-origin is simply called the *origin*; this is the set which corresponds to $\{1, \ldots, \tilde{p}\}$ before the relabeling above.

Note that the k-origin (for which there exist no shifts) is the whole set which $G_{p,k}$ permutes.

Lemma 16.2. *For a structure $\mathfrak{A}$ with the universe $\{1, \ldots, n\}$, if any of its unary relations is not constant over the origin, then the size of its orbit under $G_{p,k}$ is divisible by p^k.*

Proof. The proof is by induction on k, with the case $k = 1$ being trivial (note that $\tilde{p}$ is either p or p^2). Assume now that the lemma is known for $k - 1$. We note that either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ has a $k - 1$ origin that is different from the 1-shifted $k - 1$ origin of $\mathfrak{A}$ (because of the assumption that there exists a unary relation that is non-constant over the origin). This means that either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is not a member of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, and so size of $\mathrm{Orb}_{G_{p,k}}(\mathfrak{A})$ is larger than that of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, while both sizes are powers of p, so the lemma follows. □

Lemma 16.3. *For a structure $\mathfrak{A}$ with the universe $\{1, \ldots, n\}$, if for any of its binary relations there exists a vertex v not in the origin whose relations with the vertices in the origin are not constant, then the size of its orbit under $G_{p,k}$ is divisible by p^k.*

Proof. Again by induction, where in the case of $G_{p,1}$ we can just treat the (appropriate direction) of the relation between v and the origin as a unary relation over the origin for the purpose of bounding the orbit size. The induction step splits into three cases.

The first case is where v is outside the k-origin. In this case we just treat the relation from v to the vertices on which $G_{p,k}$ acts as a unary relation for the purpose of bounding the orbit size (remember that since the order of $G_{p,k}$ is a power of p, it is sufficient to show that the orbit size is at least p^k), and use Lemma 16.2.

The second case is where v is inside the $k - 1$-origin. In this case, similarly to the proof of Lemma 16.2, we note that either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ has a $k - 1$ origin that is different from the 1-shifted $k - 1$ origin of $\mathfrak{A}$, and so either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is different from all members of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$.

The third case is where v is in the k-origin but not in the $k - 1$-origin. Then it is in the j-shifted $k - 1$-origin for some $0 < j < \tilde{p}$. In this case either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ is such that the relations between the $k - 1$-origin and the j-shifted $k - 1$-origin are different from the relations in $\mathfrak{A}$ between the 1-shifted $k - 1$-origin and the $j + 1$-shifted (modulo $\tilde{p}$) $k - 1$-origin. Thus either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is not a member of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, showing that $\mathrm{Orb}_{G_{p,k}}(\mathfrak{A})$ is larger than $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$. □

The above lemma is the essence of what we need from this section, but the condition above for not being in a large orbit is problematic in that it is not itself closed under the action of $G_{p,k}$. However, the following corollary gives a condition that is closed under $G_{p,k}$.

Corollary 16.4. *If a binary structure over $\{1, \ldots, n\}$ has an orbit under $G_{p,k}$ whose size is not a multiple of p^k, then it is the result of a substitution of the substructures induced by its shifted (and unshifted) 1-origins in an appropriate (smaller) structure.*

Proof. For every shifted 1-origin apply Lemma 16.3 to $\sigma(\mathfrak{A})$ for an appropriate $\sigma \in G_{p,k}$; since the result of all these applications is an invariance of the relations (apart from those internal to a shifted 1-origin) under any permutation inside the shifted 1-origins, this means that the structure is the result of the appropriate substitutions (the order of substitutions is not important because they are all substitutions of different vertices of the original smaller structure). □

The above is the corollary that we will use to show a modular linear recurrence concerning structures with a finite Specker index.

16.2. Bounded Specker index implies periodicity. We now follow the method of [**67**], only instead of the permutations used there, we use the group $G_{p,k}$ to prove periodicity modulo p^k.

Let $\mathcal{C}_1, \dots, \mathcal{C}_s$ be the enumeration of all classes residing from $\mathcal{C}$. Given a sequence of integers $\mathbf{a} = (a_1, \dots, a_l)$, we define $\mathcal{C}_{\mathbf{a}}$ as the class of all structures $\mathfrak{A}$ (over our fixed language with unary and binary relations) with the universe $\{1, \dots, n\}$, such that $n \geq l$, and if one substitutes the vertex i with $\mathcal{C}_{a_i}$ for every $1 \leq i \leq l$, then one gets a structure in $\mathcal{C}$ (it is not hard to see that the order in which these substitutions are performed is not important). Note in particular that for the sequence ε of size 0 we get $\mathcal{C}_\varepsilon = \mathcal{C}$.

Claim 16.5. *If $\mathbf{a}$ is a permutation of $\mathbf{a}'$, then $d_{\mathcal{C}_{\mathbf{a}}}(n) = d_{\mathcal{C}_{\mathbf{a}'}}(n)$.*

Proof. Simple; a one to one correspondence between members of $\mathcal{C}_{\mathbf{a}}$ and the members of $\mathcal{C}_{\mathbf{a}'}$ is induced by an appropriate permutation of the vertices. □

By virtue of this claim, from now on we focus our attention on sequences $\mathbf{a}$ that are monotone nondecreasing. The ultimate periodicity results from the following two lemmas. Note that $\tilde{p}$ and $G_{p,k}$ are defined as per the preceding section.

Lemma 16.6. *If l is the length of $\mathbf{a}$ and $n \geq l + \tilde{p}^k$, then $d_{\mathcal{C}_{\mathbf{a}}}(n)$ is congruent modulo p^k to a linear sum (whose coefficients depend only on $\mathbf{a}$, p, k and $\mathcal{C}$) of functions of the type $d_{\mathcal{C}_{\mathbf{a}'}}(n - \tilde{p}^k + \tilde{p}^{k-1})$ where $\mathbf{a}'$ ranges over the sequences that are composed from $\mathbf{a}$ by inserting $\tilde{p}^{k-1}$ additional values (and in particular are of length $l + \tilde{p}^{k-1}$).*

Proof. We look at the orbits of structures in $\mathcal{C}_{\mathbf{a}}$ over $\{1, \dots, n\}$, where the permutation group G is equal to $G_{p,k}$, except that it acts on the vertices $l+1, \dots, l+\tilde{p}^k$ (instead of $1, \dots, \tilde{p}^k$). Because of Corollary 16.4, to get the number $d_{\mathcal{C}_{\mathbf{a}}}(n)$ modulo p^k we now need only concern ourselves with structures that result from substituting the substructures, induced by $l + 1 + \tilde{p}i, \dots, l + \tilde{p}(i + 1)$ for every $0 \leq i < \tilde{p}^{k-1}$, into an appropriate smaller structure.

The number of possibilities for these substitution schemes is finite (depending only on p, k and $\mathcal{C}$), and the count of the smaller structures in which the substitutions take place corresponds to the required linear combination of functions of the type $d_{\mathcal{C}_{\mathbf{a}'}}(n - \tilde{p}^k + \tilde{p}^{k-1})$. □

Lemma 16.7. *If $\mathbf{a}$ contains at least $\tilde{p}^k$ copies of the same value, then $d_{\mathcal{C}_{\mathbf{a}}}(n)$ is congruent modulo p^k to a linear sum (whose coefficients depend only on $\mathbf{a}$, p, k and $\mathcal{C}$) of functions of the type $d_{\mathcal{C}_{\mathbf{a}'}}(n - \tilde{p}^k + \tilde{p}^{k-1})$ where $\mathbf{a}'$ ranges over sequences that result from $\mathbf{a}$ by removing $\tilde{p}^k$ copies of the recurring value and inserting $\tilde{p}^{k-1}$ new values (some of which may be identical to the removed ones); in particular the length of any possible $\mathbf{a}'$ in this sum is $l - \tilde{p}^k + \tilde{p}^{k-1}$.*

Proof. Somewhat similar to the proof of Lemma 16.6. Let i be such that $a_i, \dots, a_{i+\tilde{p}^k-1}$ are identical. We look at orbits of structures in $\mathcal{C}_{\mathbf{a}}$ over $\{1, \dots, n\}$, where the permutation group G is equal to $G_{p,k}$, only that it acts on $\{i, \dots, i+\tilde{p}^k - 1\}$. Up to congruences modulo p^k we only need to look at structures that result from substituting the $\tilde{p}^{k-1}$ substructures corresponding to the shifted 1-origins of G into the appropriate structures of order $n - \tilde{p}^k + \tilde{p}^{k-1}$. □

With the above two lemmas one can complete the proof of periodicity: The set $\mathbf{A}$ of sequences $\mathbf{a}$ which do not satisfy the requirements for Lemma 16.7 is clearly finite, and it includes ε. We now define a vector function $\mathbf{f}$ from $\mathbb{N}$ to $(\mathbb{Z}_{p^k})^{|\mathbf{A}|}$ by $\mathbf{f}(n) \equiv \langle d_{\mathcal{C}_\mathbf{a}}(n) | \mathbf{a} \in \mathbf{A} \rangle \pmod{p^k}$. Using now Lemma 16.6 and Lemma 16.7 we can obtain a linear recurrence relation between $\mathbf{f}(n)$ and $\mathbf{f}(n-1), \ldots, \mathbf{f}(n-C)$, where C is bounded by $2\tilde{p}^k$ plus the size of the longest member of $\mathbf{A}$, both of which depend only on p, k, and the Specker index of $\mathcal{C}$. Since $\mathbf{f}(n)$ has a finite number of possible values for any fixed n, ultimate periodicity follows.

Acknowledgements. We are grateful to R. Bar-Yehuda, B. Courcelle, R. Roth, A. Sidi, H. Wilf and I. Yavneh for various suggestions and comments about an early draft of this paper, and to I. Gessel and M. Grohe for their comments at a later stage.

We are particularly grateful to E. Specker, for his encouragement and interest in our work, and for his various suggestions and clarifications.

References

1. *The on-line encyclopedia of integer sequences (oeis)*, http://oeis.org/.
2. M. Ajtai and R. Fagin, *Reachability is harder for directed than for undirected finite graphs*, Journal of Symbolic Logic **55.1** (1990), 113–150.
3. R. Apéry, *Irrationalité de $\zeta(2)$ et $\zeta(3)$*, Astérisque **61** (1979), 11–13.
4. József Balogh, Béla Bollobás, and Miklós Simonovits, *The number of graphs without forbidden subgraphs*, J. Comb. Theory, Ser. B **91** (2004), no. 1, 1–24.
5. József Balogh, Béla Bollobás, and David Weinreich, *The speed of hereditary properties of graphs*, J. Comb. Theory, Ser. B **79** (2000), no. 2, 131–156.
6. ______, *The penultimate rate of growth for graph properties*, Eur. J. Comb. **22** (2001), no. 3, 277–289.
7. E. Barcucci, A. Del Lungo, A Forsini, and S Rinaldi, *A technology for reverse-engineering a combinatorial problem from a rational generating function*, Advances in Applied Mathematics **26** (2001), 129–153.
8. F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, Cambridge University Press, 1998, Preprint, UQAM Université du Québec à Montréal.
9. ______, *Introduction to the theory of species of structures*, Preprint, UQAM Université du Québec à Montréal, 2008.
10. E. W. Beth, *Observations métamathématiques sur les structures simplement ordonneés*, Applications scientifiques de la logique mathématique, Collection de Logique Mathématique, Serie A, vol. 5, Paris and Louvain, 1954, pp. 29–35.
11. G.D. Birkhoff, *General theory of irregular difference equations*, Acta Mathematica **54** (1930), 205–246.
12. G.D. Birkhoff and W. J. Trjitzinsky, *Analytic theory of singular difference equations*, Acta Mathematica **60** (1933), 1–89.
13. C. Blatter and E. Specker, *Le nombre de structures finies d'une th'eorie à charactère fin*, Sciences Mathématiques, Fonds Nationale de la recherche Scientifique, Bruxelles (1981), 41–44.
14. ______, *Modular periodicity of combinatorial sequences*, Abstracts of the AMS **4** (1983), 313.
15. B. Bollobás, *Modern graph theory*, Springer, 1999.
16. M. Bousquet-Mélou, *Counting walks in the quarter plane*, Mathematics and Computer Science: Algorithms, trees, combinatorics and probabilities, Trends in Mathematics, Birkhauser, 2002, pp. 49–67.
17. D. Callan, *A combinatorial derivation of the number of labeled forests*, Journal of Integer Sequences **6** (2003), Article 03.4.7.
18. N. Chomsky and M.P. Schützenberger, *The algebraic theory of context free languages*, Computer Programming and Formal Systems (P. Brafford and D. Hirschberg, eds.), North Holland, 1963, pp. 118–161.
19. M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, *The strong perfect graph theorem*, Annals of Mathematics **164** (2006), no. 1, 51–229.

20. B. Courcelle, *The monadic second–order theory of graphs I: Recognizable sets of finite graphs*, Information and Computation **85** (1990), 12–75.
21. ———, *Monadic second order graph transductions: A survey*, Theoretical Computer Science **126** (1994), 53–75.
22. ———, *The expression of graph properties and graph transformations in monadic second-order logic*, Handbook of graph grammars and computing by graph transformations, Vol. 1 : Foundations (G. Rozenberg, ed.), World Scientific, 1997, pp. 313–400.
23. B. Courcelle and J. Engelfriet, *Graph structure and monadic second-order logic, a language theoretic approach*, Cambridge University Press, 2012, in press.
24. R. Diestel, *Graph theory*, 3 ed., Graduate Texts in Mathematics, Springer, 2005.
25. H.-D. Ebbinghaus and J. Flum, *Finite model theory*, Perspectives in Mathematical Logic, Springer, 1995.
26. ———, *Finite model theory*, Springer Verlag, 1995.
27. P. Erdös and I. Kaplansky, *The asymptotic number of latin rectangles*, Amer. J. Math. **68** (1946), 230–236.
28. G. Everest, A. van Porten, I. Shparlinski, and T. Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, vol. 104, American Mathematical Society, 2003.
29. S. Feferman, *Some recent work of Ehrenfeucht and Fraïssé*, Proceedings of the Summer Institute of Symbolic Logic, Ithaca 1957, pp. 201-209, 1957.
30. S. Feferman and R. Vaught, *The first order properties of algebraic systems*, Fundamenta Mathematicae **47** (1959), 57–103.
31. E. Fischer, *The Specker-Blatter theorem does not hold for quaternary relations*, Journal of Combinatorial Theory, Series A **103** (2003), 121–136.
32. E. Fischer and J. A. Makowsky, *The Specker-Blatter theorem revisited*, COCOON, Lecture Notes in Computer Science, vol. 2697, Springer, 2003, pp. 90–101.
33. P. Flajolet, *On congruences and continued fractions for some classical combinatorial quantities*, Discrete Mathematics **41** (1982), 145–153.
34. P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.
35. F. Gécseg and M. Steinby, *Tree languages*, Handbook of formal languages, Vol. 3 : Beyond words (G. Rozenberg and A. Salomaa, eds.), Springer Verlag, Berlin, 1997, pp. 1–68.
36. S. Gerhold, *On some non-holonomic sequences*, Electronic Journal of Combinatorics **11** (2004), 1–7.
37. A. Gertsch and A.M. Robert, *Some congruences concerning the Bell numbers*, Bull. Belg. Math. Soc. **3** (1996), 467–475.
38. I. Gessel, *Combinatorial proofs of congruences*, Enumeration and design (D.M. Jackson and S.A. Vanstone, eds.), Academic Press, 1984, pp. 157–197.
39. ———, *Counting latin rectangles*, Bull. Amer. Math. Soc. **16** (1987), 79–83.
40. ———, *Symmetric functions and p-recursiveness*, J. Comb. Theory, Ser. A **53** (1990), no. 2, 257–285.
41. F. Harary and E. Palmer, *Graphical enumeration*, Academic Press, 1973.
42. L. A. Hemaspaandra and H. Vollmer, *The satanic notations: Counting classes beyond* $\sharp\mathbf{P}$ *and other definitional adventures*, SIGACTN: SIGACT News (ACM Special Interest Group on Automata and Computability Theory) **26** (1995), 3–13.
43. W. Hodges, *Model theory*, Encyclopedia of Mathematics and its Applications, vol. 42, Cambridge University Press, 1993.
44. J. E. Hopcroft and J. D. Ullman, *Introduction to automata theory, languages and computation*, Addison-Wesley Series in Computer Science, Addison-Wesley, 1980.
45. T. Kotek and J. A. Makowsky, *Definability of combinatorial functions and their linear recurrence relations*, Fields of Logic and Computation, 2010, pp. 444–462.
46. T. Kotek and J.A. Makowsky, *A representation theorem for holonomic sequences based on counting lattice paths*, submitted, 2010.
47. ———, *A representation theorem for (q-)holonomic sequences*, submitted, 2010.
48. H. Läuchli and J. Leonhard, *On the elementary theory of order*, Fundamenta Mathematicae **59** (1966), 109–116.
49. L. Libkin, *Elements of finite model theory*, Springer, 2004.
50. E. Lucas, *Théorie des fonctions numériques simplement périodiques*, Amer. J. Math. **1** (1878), 184–240, 289–321.

51. J.A. Makowsky, *Algorithmic uses of the Feferman-Vaught theorem*, Annals of Pure and Applied Logic **126.1-3** (2004), 159–213.
52. B.D. McKay and E. Rogoyski, *Latin squares of order 10*, Electronic J. Combinatorics **2.1** (1995), N3, 1–4.
53. B.D. McKay and I.M. Wanless, *On the number of latin squares*, Ann.Comb. **9** (2005), 335–344.
54. M. Mishna, *Classifying lattice walks restricted to the quarter plane*, Journal of Combinatorial Theory, Series A **116** (2009), 460–477.
55. J.D. Monk, *Mathematical logic*, Graduate Texts in Mathematics, Springer Verlag, 1976.
56. E.M. Palmer, R.C. Read, and R.W. Robinson, *Counting claw-free cubic graphs*, Siam J. Discrete Math. **16.1** (2002), 65–73.
57. M. Petkovsek, H. Wilf, and D. Zeilberger, *A=B*, AK Peters, 1996.
58. R.C. Read, *The enumeration of locally restricted graphs, i.*, J. London Math. Soc. **34** (1959), 417–436.
59. ———, *The enumeration of locally restricted graphs, ii.*, J. London Math. Soc. **35** (1960), 344–351.
60. J.H. Redfield, *The theory of group-reduced distributions*, Amer. J. Math. **49** (1927), 433–455.
61. G.C. Rota, *The number of partitions of a set*, American Mathematical Monthly **71 (5)** (1964), 498–504.
62. U. Rotics, *Efficient algorithms for generally intractable graph problems restricted to specific classes of graphs*, Ph.D. thesis, Technion- Israel Institute of Technology, 1998.
63. G. Rozenberg and A. Salomaa (eds.), *Handbook of formal languages, vol. 3 : Beyond words*, Springer Verlag, Berlin, 1997.
64. E. R. Scheinerman and J. Zito, *On the size of hereditary classes of graphs*, J. Combin. Theory Ser. B **61** (1994), 16–39.
65. D. Seese, *The structure of the models of decidable monadic theories of graphs*, Annals of Pure Applied Logic **53** (1991), 169–195.
66. E. Specker, *Application of logic and combinatorics to enumeration problems*, Trends in Theoretical Computer Science (E. Börger, ed.), Computer Science Press, 1988, Reprinted in: Ernst Specker, Selecta, Birkhäuser 1990, pp. 324-350, pp. 141–169.
67. ———, *Modular counting and substitution of structures*, Combinatorics, Probability and Computing **14** (2005), 203–210.
68. R. P. Stanley, *Differentiably finite power series*, European Journal of Combinatorics **1** (1980), 175–188.
69. R.P. Stanley, *Enumerative combinatorics*, vol. II, Cambridge University Press, 1999.
70. D. S. Stones, *The many formulae for the number of latin rectangles*, The Electronic Journal of Combinatorics **17** (2010), A1.
71. D. S. Stones and I.M. Wanless, *Divisors of the number of latin rectangles*, J. Combin. Theory Ser. A **117** (2010), 204215.
72. L. Takács, *On the number of distinct forests*, SIAM J. Disc. Math. **3.4** (1990), 574–581.
73. H.S. Wilf, *generatingfunctionology*, Academic Press, 1990.
74. N.C. Wormald, *The number of labelled cubic graphs with no triangles*, Proceedings of the 66 Twelfth Southeastern Conference on Combinatorics, Graph Theory, and Computing, vol. 2, 1981, pp. 359–378.
75. D. Zeilberger, *A holonomic systems approach to special functions identities*, J. of Computational and Applied Mathematics **32** (1990), 321–368.

Faculty of Computer Science, Technion–Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: `eldar@cs.technion.ac.il`

Faculty of Computer Science, Technion–Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: `tkotek@cs.technion.ac.il`

Faculty of Computer Science, Technion–Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: `janos@cs.technion.ac.il`