Application of Logic to Integer Sequences: A Survey

Johann A. Makowsky*

Department of Computer Science
Technion–Israel Institute of Technology, Haifa, Israel
janos@cs.technion.ac.il

Abstract. Chomsky and Schützenberger showed in 1963 that the sequence $d_L(n)$, which counts the number of words of a given length n in a regular language L, satisfies a linear recurrence relation with constant coefficients for n, or equivalently, the generating function $g_L(x) = \sum_n d_L(n)x^n$ is a rational function. In this talk we survey results concerning sequences $a(n)$ of natural numbers which

- satisfy linear recurrence relations over $\mathbb{Z}$ or $\mathbb{Z}_m$, and
- have a combinatorial or logical interpretation.

We present 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 recent results by T. Kotek and the author.

For Ernst Specker on the occasion of his 90th birthday

1 Sequences of Integers and Their Combinatorial Interpretations

In this talk we discuss sequences $a(n)$ of natural numbers or integers which arise in combinatorics. Many such sequences satisfy linear recurrence relations with constant or polynomial coefficients. The traditional approach to the study of 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. [FS09]. There is a substantial theory of how to verify and prove identities among the terms of $a(n)$, see [PWZ96].

We are interested in the case where $a(n)$ admits a *combinatorial* or a *logical* interpretation, i.e., $a(n)$ counts the number of some relations or functions on the set $[n] = \{1, \ldots, n\}$ which have a certain property possibly definable in some logical formalism (with or without its natural order). To make this precise, we assume the audience is familiar with the very basics of Logic and Finite

* Partially supported by a grant of the Fund for Promotion of Research of the Technion–Israel Institute of Technology and grant ISF 1392/07 of the Israel Science Foundation (2007-2010).

A. Dawar and R. de Queiroz (Eds.): WoLLIC 2010, LNAI 6188, pp. 34–41, 2010.

Species

Model Theory, cf. [EF95, Lib04]. A more general framework for combinatorial interpretations of counting functions is described in [BLL98]. Lack of time does not allow us to use this formalism in this talk. 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 on a set of n elements equipped with unary predicates, cf. [EF95].

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

Definition 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* $\tau = \{R_1, \ldots R_r\} = \{\bar{R}\}$ *or* $\tau_{ord} = \{<_{nat}, \bar{R}\}$ *with finite universe* $[n] = \{1, \ldots, n\}$ *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*

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

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.*

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 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* **SOL**(τ_{ord}) *(***MSOL**(τ_{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* **SOL**[1]*.*

Remarks 1

(i) *If $a(n)$ has a combinatorial interpretation then for all $n \in \mathbb{N}$ we have* $a(n) \geq 0$.

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

(ii) *There are only countably many Specker sequences.*
(iii) *Every Specker sequence is computable, and in fact it is in* $\sharp \cdot \mathbf{PH}$*, [HV95], hence computable in exponential time.*
(iv) *The set of Specker sequences is closed under the point-wise operations of addition and multiplication. The same holds for* **MSOL**-*Specker sequences.*

2 Linear Recurrences

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

Definition 3 (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) MC-finite *(modularly C-finite), if for every* $m \in \mathbb{N}, m > 0$ *there is* $q(m) \in \mathbb{N} \backslash \{0\}$ *for which* $a(n)$ *satisfies for all* $n > q(m)$

Is equivalence decidable for MC-finite sequences?

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

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.*

(iv) 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* $q = 1$.

The terminology C-finite and holonomic are due to [Zei90]. P-recursive is due to [Sta80]. P-recursive sequences were already studied in [Bir30, BT33].

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

Lemma 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 $\mid a(n) \mid \leq n!^{\gamma}$ 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 integer m, [Ger04].

Proposition 1. *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.

Proposition 2

(i) *There are only countably many P-recursive sequences $a(n)$.*
(ii) *There are continuum many MC-finite sequences.*

3 Logical Interpretations and Linear Recurrences

Modular recurrence relations for sequences with combinatorial interpretation are studied widely, cf. [Fla82, Ges84]. A logical approach to this topic was pioneered in [BS81, BS83] and further pursued in [Spe88, Spe05]. C. Blatter and E. Specker have shown:

Theorem 1 (C. Blatter and E. Specker, [BS81]). *Let $a(n)$ be a Specker sequence which has a* pure **MSOL**-*interpretation $\mathcal{K}$ over a finite vocabulary which contains only relation symbols of arity at most two. Then $a(n)$ is MC-finite.*

Remarks 2

(i) *Theorem 1 is not true for* **MSOL**-*interpretations with order, i.e. which are not pure, cf. [FM03].*
(ii) *E. Fischer, [Fis03], showed that it is also not true if one allows relation symbols of arity ≥ 4, see also [Spe05].*
(iii) *In the light of Remark 1(ii) and Proposition 2(ii) there cannot be a converse of Theorem 1.*

In 1984 I. Gessel proved the following related result:

Theorem 2 (I. Gessel, [Ges84]). *Let $\mathcal{K}$ be a class of (possibly) directed graphs of bounded degree d which is closed under disjoint unions and components. Then $d_{\mathcal{K}}(n)$ is MC-finite.*

Remark 3. *Theorem 2 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. Let us call a class of directed graphs $\mathcal{K}$ a* Gessel class *if $\mathcal{K}$ is closed under disjoint unions and components and its members are of bounded degree. 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. [EF95]. Inspired by Theorem 1 and Theorem 2, E. Fischer and the author showed:

Theorem 3 (E. Fischer and J.A. Makowsky, [FM03]). *Let $a(n)$ be a Specker sequence which has a* pure **MSOL***-interpretation $\mathcal{K}$ over any finite relational vocabulary (without restrictions on the arity of the relation symbols), but which is of bounded degree. Then $a(n)$ is MC-finite.*

Let $\mathcal{K}$ be a combinatorial or logical interpretation of $a(n)$. In [Spe88] E. Specker asks whether one can formulate a definability condition on $\mathcal{K}$ which ensures that $a(n)$ is SP-recursive. There are really two questions here:

Question A: Can one formulate a definability condition on combinatorial interpretations $\mathcal{K}$ of $a(n)$ which ensures that $a(n)$ is SP-recursive.

Question B: Can one formulate a definability condition on **pure** combinatorial interpretations $\mathcal{K}$ of $a(n)$ which ensures that $a(n)$ is SP-recursive.

We shall see that the answer to Question A is in the affirmative, but that Question B remains open.

We first note that for C-finite sequences the answer to Question A is affirmative.

Theorem 4 (N. Chomsky and M. Schützenberger, [CS63]). *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 ([KM09]). *Let $a(n)$ be a function $a : \mathbb{N} \to \mathbb{Z}$ 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 $a(n) = d_1(n) - d_2(n)$.*

Remark 4. *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 well-known characterization of regular languages in **MSOL**, Theorem 4 and Theorem 5 can be combined, using Lemma 1.

Theorem 6. *Let $a(n)$ be a function $a : \mathbb{N} \to \mathbb{Z}$. $a(n)$ 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 $a(n) = d_1(n) - d_2(n)$.*

4 P-Recursive (Holonomic) Sequences

In the final part of the talk we answer E. Specker's Question A positively by giving two a characterization of P-recursive sequences both inspired by Theorem 6. We also discuss why Question B seems harder to answer.

Both characterizations involve regular languages L over an alphabet Σ, or equivalently, both use **MSOL**-interpretations.

In the first characterization, *regular languages* are augmented by a set of legal *Lattice Paths*, and are called LP-interpretations and have no weights. More precisely, we count not only words in L, but the words together with functions which map positions of the word w of length n into $[n]$ subject to certain mild restrictions. The graphs of these functions are reminiscent of lattice paths, [GJ83, GR96].

In the second characterization, *regular languages* are equipped with *weights* which depend both on the letter in the word, and the position of this letter. They are called PW-interpretations. More precisely, we count weighted words in a language L where the weight is defined by a position specific scoring matrix, widely used in computational biology to search DNA and protein databases for sequence similarities, cf. [SSGE82, AMS+97]. This approach is also reminiscent to counting weighted homomorphisms, cf. [BCL+06]. Position-independent weights on words were used in [NZ99] for the extension of the powerful (and so far under-utilized) Goulden-Jackson Cluster method for finding the generating function for the number of words avoiding, as factors, the members of a prescribed set. In [Ges84] position-independent weights are used to prove modular congruences.

Acknowledgments

All the new results in this survey are taken from T. Kotek's ongoing work on his Ph.D. thesis [KM10b, KM10a]. I would like to thank T. Kotek for allowing me to use entire passages from our joint manuscripts in this extended abstract of my invited lecture.

References

[AMS+97] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped blast and psi-blast: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402 (1997)

[BCL+06] Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Counting graph homomorphisms. In: Klazar, M., Kratochvil, J., Loebl, M., Matousek, J., Thomas, R., Valtr, P. (eds.) Topics in Discret mathematics, pp. 315–371. Springer, Heidelberg (2006)

[Bir30] Birkhoff, G.D.: General theory of irregular difference equations. Acta Mathematica 54, 205–246 (1930)

[BLL98] Bergeron, F., Labelle, G., Leroux, P.: Combinatorial Species and Tree-like Structures. Encyclopedia of Mathematics and its Applications, vol. 67. Cambridge University Press, Cambridge (1998)

[BS81] Blatter, C., Specker, E.: Le nombre de structures finies d'une th'eorie à charactère fin. Sciences Mathématiques, Fonds Nationale de la recherche Scientifique, Bruxelles, 41–44 (1981)

[BS83] Blatter, C., Specker, E.: Modular periodicity of combinatorial sequences. Abstracts of the AMS 4, 313 (1983)

[BT33] Birkhoff, G.D., Trjitzinsky, W.J.: Analytic theory of singular difference equations. Acta Mathematica 60, 1–89 (1933)

[CS63] Chomsky, N., Schützenberger, M.P.: The algebraic theory of context free languages. In: Brafford, P., Hirschberg, D. (eds.) Computer Programming and Formal Systems, pp. 118–161. North Holland, Amsterdam (1963)

[EF95] Ebbinghaus, H.D., Flum, J.: Finite Model Theory. In: Perspectives in Mathematical Logic, Springer, Heidelberg (1995)

[EvPSW03] Everest, G., van Porten, A., Shparlinski, I., Ward, T.: Recurrence Sequences. Mathematical Surveys and Monographs, vol. 104. American Mathematical Society, Providence (2003)

[Fis03] Fischer, E.: The Specker-Blatter theorem does not hold for quaternary relations. Journal of Combinatorial Theory, Series A 103, 121–136 (2003)

[Fla82] Flajolet, P.: On congruences and continued fractions for some classical combinatorial quantities. Discrete Mathematics 41, 145–153 (1982)

[FM03] Fischer, E., Makowsky, J.A.: The Specker-Blatter theorem revisited. In: Warnow, T.J., Zhu, B. (eds.) COCOON 2003. LNCS, vol. 2697, pp. 90–101. Springer, Heidelberg (2003)

[FS09] Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press, Cambridge (2009)

[Ger04] Gerhold, S.: On some non-holonomic sequences. Electronic Journal of Combinatorics 11, 1–7 (2004)

[Ges84] Gessel, I.: Combinatorial proofs of congruences. In: Jackson, D.M., Vanstone, S.A. (eds.) Enumeration and design, pp. 157–197. Academic Press, London (1984)

[GJ83] Goulden, I.P., Jackson, D.M.: Combinatorial Enumeration. Interscience Series in Discrete Mathematics. Wiley, Chichester (1983)

[GR96] Gessel, I.M., Ree, S.: Lattice paths and Faber polynomials. In: Balakrishnan, N. (ed.) Advances in combinatorial methods and applications to probability and statistics, pp. 3–14. Birkhäuser, Basel (1996)

[HV95] Hemaspaandra, V.: 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)

[KM09] Kotek, T., Makowsky, J.A.: Definability of combinatorial functions and their linear recurrence relations. Electronically available at arXiv:0907.5420 (2009)

[KM10a] Kotek, T., Makowsky, J.A.: Application of logic to generating functions: Holonomic sequences. Manuscript (2010)

[KM10b] Kotek, T., Makowsky, J.A.: A representation theorem for holonomic sequences. Manuscript (2010)

[Lib04] Libkin, L.: Elements of Finite Model Theory. Springer, Heidelberg (2004)

[NZ99] Noonan, J., Zeilberger, D.: The Goulden-Jackson cluster method: Extensions, applications and implementations. J. Differ. Equations Appl. 5(4-5), 355–377 (1999)

[PWZ96] Petkovsek, M., Wilf, H., Zeilberger, D.: A=B. AK Peters, Wellesley (1996)

[Spe88] Specker, E.: Application of logic and combinatorics to enumeration problems. In: Börger, E. (ed.) Trends in Theoretical Computer Science, pp. 141–169. Computer Science Press, Rockville (1988); Reprinted in: Ernst Specker, Selecta, Birkhäuser 1990, pp. 324–350

[Spe05] Specker, E.: Modular counting and substitution of structures. Combinatorics, Probability and Computing 14, 203–210 (2005)

[SSGE82] Stormo, G.D., Schneider, T.D., Gold, L., Ehrenfeucht, A.: Use of the 'perceptron' algorithm to distinguish translational initiation sites in e. coli. Nucleic Acid Research 10, 2997–3012 (1982)

[Sta80] Stanley, R.P.: Differentiably finite power series. European Journal of Combinatorics 1, 175–188 (1980)

[Zei90] Zeilberger, D.: A holonomic systems approach to special functions identities. J. of Computational and Applied Mathematics 32, 321–368 (1990)