From Equational Specifications of Algebras with Structure to Varieties of Data Languages

Stefan Milius
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
mail@stefan-milius.eu

Abstract

This extended abstract first presents a new category theoretic approach to equationally axiomatizable classes of algebras. This approach is well-suited for the treatment of algebras equipped with additional computationally relevant structure, such as ordered algebras, continuous algebras, quantitative algebras, nominal algebras, or profinite algebras. We present a generic HSP theorem and a sound and complete equational logic, which encompass numerous flavors of equational axiomatizations studied in the literature. In addition, we use the generic HSP theorem as a key ingredient to obtain Eilenberg-type correspondences yielding algebraic characterizations of properties of regular machine behaviours. When instantiated for orbit-finite nominal monoids, the generic HSP theorem yields a crucial step for the proof of the first Eilenberg-type variety theorem for data languages.

2012 ACM Subject Classification Theory of computation → Equational logic and rewriting; Theory of computation → Formal languages and automata theory

Keywords and phrases Birkhoff theorem, Equational logic, Eilenberg theorem, Data languages

Digital Object Identifier 10.4230/LIPIcs.CALCO.2019.2

Category Invited Paper

Funding Supported by Deutsche Forschungsgemeinschaft (DFG) under project MI 717/5-2.

1 Equations and Algebras with Structure

A key tool in the algebraic theory of data structures is their specification by operations (constructors) and equations that they ought to satisfy. Birkhoff's celebrated HSP theorem [7] states that a class of algebras over a signature Σ is a *variety* (i.e. closed under homomorphic images, subalgebras, and products) iff it is axiomatizable by equations $s = t$ between Σ-terms. Birkhoff also introduced a complete deduction system for reasoning about equations.

In algebraic approaches to the semantics of programming languages and computational effects, it is often natural to study algebras whose underlying sets are equipped with additional computationally relevant structure and whose operations preserve that structure. An important line of research thus concerns extensions of Birkhoff's theory of equational axiomatization beyond ordinary Σ-algebras. On the syntactic level, this requires to enrich Birkhoff's notion of an equation in ways that reflect the extra structure. For example, Bloom [8] and Adámek et al. [1, 2] established versions of the HSP theorem for *ordered algebras* and *continuous* ones, respectively, along with complete deduction systems. Here, the role of equations $s = t$ is taken over by inequations $s \leq t$. Recently, Mardare, Panangaden and Plotkin [19, 20] presented an HSP theorem for *quantitative algebras* and a complete deduction system. In the quantitative setting, equations $s =_\epsilon t$ are equipped with a non-negative real number ϵ, interpreted as "s and t have distance at most ϵ". Varieties of *nominal algebras* were studied by Gabbay [15] and Kurz and Petrişan [18]. Here, the appropriate syntactic concept involves equations $s = t$ with constraints on the support of their variables. Finally, Reiterman [29] as well as Eilenberg and Schützenberger [13] showed that *pseudovarieties* (i.e. classes of *finite* algebras closed under homomorphic images, subalgebras and *finite*

© Stefan Milius;
licensed under Creative Commons License CC-BY
8th Conference on Algebra and Coalgebra in Computer Science (CALCO 2019).
Editors: Markus Roggenbach and Ana Sokolova; Article No. 2; pp. 2:1–2:5
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

products) can be axiomatized by so-called *profinite equations* or, equivalently, by sequences of ordinary equations $(s_i = t_i)_{i<\omega}$, interpreted as "all but finitely many of the equations $s_i = t_i$ hold".

We propose a general category theoretic framework that allows to study equationally specified classes of algebras with extra structure in a systematic way. Our overall goal is to isolate the domain-specific part of any theory of equational axiomatization from its generic core. Our framework is parametric in the following data:

- a category $\mathscr{A}$ with a factorization system $(\mathscr{E}, \mathscr{M})$;
- a full subcategory $\mathscr{A}_0 \subseteq \mathscr{A}$;
- a class Λ of cardinal numbers;
- a class $\mathscr{X} \subseteq \mathscr{A}$ of objects.

Here, $\mathscr{A}$ is the category of algebras under consideration (e.g. ordered algebras, quantitative algebras, nominal algebras). Varieties are formed within $\mathscr{A}_0$, and the cardinal numbers in Λ determine the arities of products under which the varieties are closed. Thus, the choice $\mathscr{A}_0 =$ finite algebras and $\Lambda =$ finite cardinals corresponds to pseudovarieties, and $\mathscr{A}_0 = \mathscr{A}$ and $\Lambda =$ all cardinals to varieties. The crucial ingredient of our setting is the parameter $\mathscr{X}$, which is the class of objects over which equations are formed. Typically, $\mathscr{X}$ is chosen to be some class of freely generated algebras in $\mathscr{A}$. Equations are modeled as $\mathscr{E}$-quotients $e\colon X \twoheadrightarrow E$ (more generally, filters of such quotients) with domain $X \in \mathscr{X}$.

The choice of $\mathscr{X}$ reflects the desired expressivity of equations in a given setting, and it determines the type of quotients under which equationally axiomatizable classes are closed. More precisely, in our category theoretic framework a *variety* is defined to be a subclass of $\mathscr{A}_0$ closed under $\mathscr{E}_{\mathscr{X}}$-quotients, $\mathscr{M}$-subobjects, and Λ-products, where $\mathscr{E}_{\mathscr{X}}$ is a subclass of $\mathscr{E}$ derived from $\mathscr{X}$. Due to its parametric nature, this concept of a variety is widely applicable and turns out to specialize to many interesting cases. The main result is the

▶ **General HSP Theorem** [22]. *A subclass of $\mathscr{A}_0$ forms a variety if and only if it is axiomatizable by equations.*

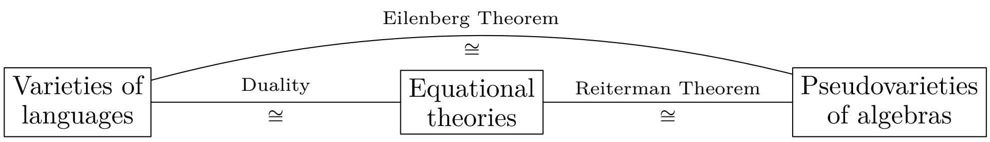
In addition, we introduce a generic deduction system for equations, based on two simple proof rules for equations $e\colon X \twoheadrightarrow E$, and establish a

▶ **General Completeness Theorem** [22]. *The generic deduction system for equations is sound and complete.*

The above two theorems can be seen as the generic building blocks of the model theory of algebras with structure. They form the common core of numerous Birkhoff-type results and give rise to a systematic recipe for deriving concrete HSP and completeness theorems.

2 Varieties of Data Languages

Since the above results also cover Reiterman-type results (via the choice of $\mathscr{X}$ as free algebras over *finite* sets) the General HSP Theorem yields a key tool for a generic algebraic language theory. In this theory one studies formal languages and other types of behaviours of finite machines (e.g. weighted languages, infinite words, trees, cost functions) in terms of algebraic structures that recognize them. As a prime example, regular languages can be described purely algebraically as the languages recognized by finite monoids, and a celebrated result by McNaughton, Papert, and Schützenberger [21, 33] asserts that a regular language is definable in first-order logic if and only if its syntactic monoid is aperiodic (i.e. it satisfies the equation $x^{n+1} = x^n$ for sufficiently large n). As an immediate application, this algebraic characterization yields an effective procedure for deciding first-order definability. The first systematic approach to correspondence results of this kind was initiated by Eilenberg [14] who proved that *varieties of languages* (i.e. classes of regular languages closed under the set-theoretic boolean operations, derivatives, and homomorphic preimages) correspond bijectively to pseudovarieties of monoids. Inspired by Eilenberg's work, over the past four decades numerous further variety theorems were discovered for regular languages [16, 24, 27, 36], treating notions of varieties with modified closure properties, but also for machine behaviors beyond finite words, including weighted languages over a field [30], infinite words [25, 38], words on linear orderings [5, 6], ranked trees [4], binary trees [32], and cost functions [12]. Recent research [3, 9] has focused on generic approaches and has culminated in Salamanca's work [31] and our General Eilenberg Theorem that covers all of the above ones as special instances [37]. Its proof is based on two key ingredients: (1) *duality* in order to establish a correspondence between (profinite) equational theories and varieties of recognizable languages and (2) a generic *Reiterman-type correspondence* to pseudovarieties.



That duality plays an important role for Eilenberg-type correspondences has been established by Gehrke, Grigorieff and Pin [16]. The duality based proofs in [31, 37] yield a blueprint for new correspondences of this kind. For example, it allows to obtain the first Eilenberg-type correspondence for *data languages* [22]. Such languages are of significant interest in recent years, driven by practical applications in various areas of computer science, including efficient parsing of XML documents or software verification. Mathematically, data languages are modeled using *nominal sets* (see e.g. [26]). Over the years, several machine models for handling data languages of different expressive power have been proposed; see [34, 35] for a comprehensive survey. Here we focus on languages recognized by *orbit-finite nominal monoids*. They form an important subclass of the languages accepted by Francez and Kaminski's *finite memory automata* [17] (which are expressively equivalent to orbit-finite automata in the category of nominal sets [11]) and have been characterized by a fragment of monadic second-order logic over data words called *rigidly guarded MSO* [28]. In addition, Bojańczyk [10] and Colcombet, Ley and Puppis [28] established nominal versions of the McNaughton-Papert-Schützenberger theorem and showed that the first-order definable data languages are precisely the ones recognizable by aperiodic orbit-finite monoids. It is therefore natural to ask whether an Eilenberg-type theorem can be developed for data languages, and we answer this question affirmatively:

▶ **Nominal Eilenberg Theorem** [23]**.** *Varieties of data languages correspond bijectively to pseudovarieties of nominal monoids.*

Here, the notion of a *pseudovariety of nominal monoids* is as expected: a class of orbit-finite nominal monoids closed under quotient monoids, submonoids, and finite products. In contrast, the notion of a *variety of data languages* requires two extra conditions unfamiliar from other Eilenberg-type correspondences, most notably a technical condition called *completeness.* Like the original Eilenberg theorem, its nominal version gives rise to a generic relation between properties of data languages and properties of nominal monoids. For instance, the aperiodic orbit-finite monoids form a pseudovariety, and the first-order definable data languages form a variety, and thus the equivalence of these concepts can be understood as an instance of the nominal Eilenberg correspondence.

It should be pointed out that the Nominal Eilenberg Theorem requires new techniques and cannot be obtained as a mere instance of the previous General Eilenberg Theorem, since the latter is based on working with algebraic-like base categories (which excludes nominal sets) and the recognition by finite structures. However, our approach can be seen as an indication of the robustness of the key ideas behind the duality-based methodology for algebraic recognition and the guidance they provide towards future applications and results.

References

1. J. Adámek, A. H. Mekler, E. Nelson, and J. Reiterman. On the logic of continuous algebras. *Notre Dame J. Formal Logic*, 29(3):365–380, 1988.
2. Jiři Adámek, Evelyn Nelson, and Jan Reiterman. The Birkhoff Variety Theorem for continuous algebras. *Algebra Universalis*, 20(3):328–350, 1985.
3. Jiří Adámek, Stefan Milius, Robert S.R. Myers, and Henning Urbat. Generalized Eilenberg Theorem: Varieties of Languages in a Category. *ACM Trans. Comput. Log.*, 20(1):3:1–3:47, 2019.
4. J. Almeida. On pseudovarieties, varieties of languages, filters of congruences, pseudoidentities and related topics. *Algebra Universalis*, 27(3):333–350, 1990.
5. N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. In *Proc. LATIN'98*, volume 1380 of *LNCS*, pages 53–64. Springer, 1998.
6. N. Bedon and C. Rispal. Schützenberger and Eilenberg theorems for words on linear orderings. In *Proc. DLT'05*, volume 3572 of *LNCS*, pages 134–145. Springer, 2005.
7. Garret Birkhoff. On the structure of abstract algebras. *Proceedings of the Cambridge Philosophical Society*, 10:433—-454, 1935.
8. Stephen L. Bloom. Varieties of ordered algebras. *J. Comput. Syst. Sci.*, 2(13):200–212, 1976.
9. M. Bojańczyk. Recognisable languages over monads. In I. Potapov, editor, *Proc. DLT'15*, volume 9168 of *LNCS*, pages 1–13. Springer, 2015. `arXiv:1502.04898`.
10. Mikołaj Bojańczyk. Nominal Monoids. *Theory of Computing Systems*, 53(2):194–222, 2013.
11. Mikołaj Bojańczyk, Bartek Klin, and Sławomir Lasota. Automata theory in nominal sets. *Log. Methods Comput. Sci.*, 10(3:4):44 pp., 2014.
12. L. Daviaud, D. Kuperberg, and J.-É. Pin. Varieties of Cost Functions. In N. Ollinger and H. Vollmer, editors, *Proc. STACS 2016*, volume 47 of *LIPIcs*, pages 30:1–30:14. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016.
13. S. Eilenberg and M. P. Schützenberger. On pseudovarieties. *Advances Math.*, 10:413–418, 1976.
14. Samuel Eilenberg. *Automata, Languages, and Machines Vol. B.* Academic Press, 1976.
15. Murdoch J. Gabbay. Nominal algebra and the HSP theorem. *Journal of Logic and Computation*, 19:341–367, 2009.
16. Mai Gehrke, Serge Grigorieff, and Jean-Éric Pin. Duality and equational theory of regular languages. In *Proc. ICALP'08, Part II*, volume 5126 of *LNCS*, pages 246–257. Springer, 2008.
17. Michael Kaminski and Nissim Francez. Finite-memory automata. *Theoret. Comput. Sci.*, 134(2):329–363, 1994.
18. Alexander Kurz and Daniela Petrisan. On universal algebra over nominal sets. *Mathematical Structures in Computer Science*, 20(2):285–318, 2010.
19. Radu Mardare, Prakash Panangaden, and Gordon Plotkin. Quantitative Algebraic Reasoning. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '16, pages 700–709. ACM, 2016.
20. Radu Mardare, Prakash Panangaden, and Gordon Plotkin. On the axiomatizability of quantitative algebras. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12. IEEE Computer Society, 2017. `doi:10.1109/LICS.2017.8005102`.

21 Robert McNaughton and Seymour A. Papert. *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971.
22 Stefan Milius and Henning Urbat. Equational Axiomatization of Algebras with Structure. In Mikołaj Bojańczyk and Alex Simpson, editors, *Proc. Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 11425 of *Lecture Notes Comput. Sci. (ARCoSS)*, pages 400–417, 2019.
23 Stefan Milius and Henning Urbat. Varietes of Data Languages. In *Proc. 46th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 132 of *LIPIcs*, pages 130:1–130:14, 2019. (Full version available online at `arXiv:1903.08053`). `doi:10.4230/LIPIcs.ICALP.2019.130`.
24 J.-É. Pin. A variety theorem without complementation. *Russ. Math.*, 39:80–90, 1995.
25 J.-É. Pin. Positive varieties and infinite words. In *LATIN 98*, volume 1380 of *LNCS*, pages 76–87. Springer, 1998.
26 Andrew M. Pitts. *Nominal Sets: Names and Symmetry in Computer Science*. Cambridge University Press, 2013.
27 L. Polák. Syntactic semiring of a language. In J. Sgall, A. Pultr, and P. Kolman, editors, *Proc. MFCS'01*, volume 2136 of *LNCS*, pages 611–620. Springer, 2001.
28 Gabriele Puppis, Thomas Colcombet, and Clemens Ley. Logics with rigidly guarded data tests. *Log. Methods Comput. Sci.*, 11(3:10):56 pp., 2015.
29 J. Reiterman. The Birkhoff theorem for finite algebras. *Algebra Universalis*, 14(1):1–10, 1982.
30 C. Reutenauer. Séries formelles et algèbres syntactiques. *J. Algebra*, 66:448–483, 1980.
31 Julian Salamanca. Unveiling Eilenberg-type Correspondences: Birkhoff's Theorem for (finite) Algebras + Duality, February 2017. `arXiv:1702.02822`.
32 S. Salehi and M. Steinby. Tree algebras and varieties of tree languages. *Theor. Comput. Sci.*, 377(1-3):1–24, 2007.
33 Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Inform. and Control*, 8:190–194, 1965.
34 Thomas Schwentick. Automata for XML – A survey. *J. Comput. System Sci.*, 73(3):289–315, 2007.
35 Luc Segoufin. Automata and Logics for Words and Trees over an Infinite Alphabet. In *Computer Science Logic, 20th International Workshop, CSL 2006, 15th Annual Conference of the EACSL, Szeged, Hungary, September 25-29, 2006, Proceedings*, pages 41–57, 2006.
36 H. Straubing. On logical descriptions of regular languages. In S. Rajsbaum, editor, *LATIN 2002 Theor. Informatics*, volume 2286 of *LNCS*, pages 528–538. Springer, 2002.
37 Henning Urbat, Jiří Adámek, Liang-Ting Chen, and Stefan Milius. Eilenberg Theorems for Free. In Kim G. Larsen, Hans L. Bodlaender, and Jean-François Raskin, editors, *Proc. 42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *LIPIcs*. Schloss Dagstuhl, 2017. EATCS Best Paper Award.
38 T. Wilke. An Eilenberg Theorem for ∞-Languages. In *Proc. ICALP'91*, volume 510 of *LNCS*, pages 588–599. Springer, 1991.