A Fully Equational Proof of Parikh's Theorem

Luca Aceto Zoltán Ésik* Anna Ingólfsdóttir
BRICS†
Department of Computer Science
University of Aalborg
Fredrik Bajers Vej 7E
DK-9220 Aalborg

Abstract

We show that the validity of Parikh's theorem for context-free languages depends only on a few equational properties of least pre-fixed points. Moreover, we exhibit an infinite basis of μ-term equations of continuous commutative idempotent semirings.

AMS Subject Classification (2000): 03C05, 16Y60, 68Q42, 68Q45, 68Q70.
Keywords and Phrases: Parikh's theorem, commutative context-free languages, commutative rational languages, equational logic, varieties, complete axiomatizations, commutative idempotent semirings, algebraically complete commutative idempotent semirings.

A classic result of the theory of context-free languages is Parikh's theorem [25] that asserts that the letter occurrence vectors (Parikh vectors) corresponding to the words of a context-free language on a k-letter alphabet form a semilinear subset of $\mathbb{N}^k$, the free commutative monoid of k-dimensional vectors over the naturals. The theorem is usually proved by combinatorial arguments on the derivation trees of the context-free grammar, and is regarded as one of the most fundamental, yet subtly difficult to prove, in the theory of context-free languages. However, as Pilling [27] observed, Parikh's theorem may be formulated as an assertion about "rational functions" on the (free) continuous commutative idempotent semiring of all subsets of $\mathbb{N}^k$. Subsequently, Kuich [20] generalized Parikh's result to all continuous commutative idempotent semirings

*Permanent address: Dept. of Computer Science, University of Szeged, P.O.B. 652, 6701 Szeged, Hungary. Supported in part by the National Foundation of Hungary for Scientific Research.

†**B**asic **R**esearch **i**n **C**omputer **S**cience, Centre of the Danish National Research Foundation.

(l-semirings). (See also [30] for a related treatment.) In fact, by introducing rational terms that denote rational functions, or more generally, recursion terms or μ-terms denoting functions that arise as least solutions of systems of polynomial fixed point equations, Parikh's theorem can be translated into a statement about the equational theory of continuous commutative idempotent semirings:

> *For every μ-term t there exists a rational term r such that the equation $t = r$ holds in all continuous commutative idempotent semirings.*

Alternatively, one may just consider rational terms and prove that for each rational term $t(x, y_1, \ldots, y_n)$ in the variables $x, y_1, \ldots, y_n$, there is a rational term $r(y_1, \ldots, y_n)$ containing no occurrence of x that provides least solution to the fixed point equation $x = t(x, y_1, \ldots, y_n)$ over *all* continuous commutative idempotent semirings. This approach has been pursued by Hopkins and Kozen in [14], in their argument lifting Parikh's theorem to all commutative idempotent semirings with enough least fixed points to provide solutions to recursion equations. By proving this more general result, Hopkins and Kozen have shown how to replace the analytic arguments of Pilling and Kuich by arguments based only on the least (pre-)fixed point rule (also known as the Park induction rule [26]), the fixed point equation, and the algebraic laws of the sum and product operations. But since Parikh's theorem is a claim about equational theories, one would eventually like to have a fully equational proof of it. In this paper, we derive Parikh's theorem from a small set of purely equational axioms involving fixed points.

Parikh's theorem is not the only result of automata and language theory that can be derived by simple *equational reasoning* from the algebraic properties of fixed points. Other applications of the equational logic of fixed points include proofs of Kleene's theorem and its generalizations [4] (see also [20, 21, 5], where the presentation is not fully based on equational reasoning), and of Greibach's normal form theorem for context-free grammars [9]. The methods employed in the papers [19, 8] even indicate that one can embed the proof of the Krohn-Rhodes decomposition theorem [11] for finite automata and semigroups within equational logic. Further applications of fixed point theory include an equational proof of the soundness and relative completeness of Hoare's logic [2, 3]. (See also [18], and [23] for a not fully equational treatment.)

We will consider *terms*, or *μ-terms*, defined by the following (abstract) syntax, where x ranges over a fixed countable set X of variables:

$$T \quad ::= \quad x \mid T + T \mid T \cdot T \mid 0 \mid 1 \mid \mu x.T \ .$$

Thus, if t is a term and x is a variable, then $\mu x.t$ is a term. Such terms t may be interpreted as continuous functions $t_S : S^X \to S$ over continuous semirings S, and continuous idempotent semirings in particular. Such a semiring $S = (S, +, \cdot, 0, 1)$ has an idempotent additive structure, and equipped with the order $\leq$ induced by the additive structure it has all suprema. Moreover, the $\cdot$

operation preserves all suprema. (It is clear that the $+$ operation also preserves all suprema.) A prime example of such a semiring is the semiring L_{Σ^*} of all languages in Σ^*, where Σ is a finite or infinite alphabet, equipped with the union and concatenation operations as sum and product, and the empty set and the set whose unique element is the empty word as 0 and 1, respectively. In fact, one can easily show that the semiring L_{Σ^*} is freely generated by Σ in the category of continuous idempotent semirings and continuous semiring homomorphisms. More generally, if M is any monoid, then $P(M)$, the set of all subsets of M, equipped with the operations of set union and complex product as sum and product is continuous. In continuous semirings, terms of the form $\mu x.t$ are interpreted by least (pre-)fixed points.

As the semirings L_{Σ^*} are the free continuous idempotent semirings, an equation $t = t'$ between μ-terms with free variables in $Y = \{y_1, \ldots, y_n\}$ holds in all continuous idempotent semirings if and only if $t_{L_{Y^*}}(\{y_1\}, \ldots, \{y_n\}) = t'_{L_{Y^*}}(\{y_1\}, \ldots, \{y_n\})$, i.e., when t and t', viewed as context-free grammars, generate the same language in Y^*. Since the equivalence of context-free grammars is undecidable (see, e.g., [13]), it follows that, for μ-terms, the equational theory of continuous idempotent semirings is undecidable. Moreover, since the inequivalence of context free grammars is semidecidable, the equational theory of continuous idempotent semirings is not recursively enumerable. Thus, the equational theory of continuous semirings has no recursively enumerable basis. The same holds for continuous semirings, since the free continuous semirings are the semirings of power series with coefficients in the semiring of naturals equipped with a top element, and since the equality of such algebraic series is also not semidecidable.

For continuous *commutative* idempotent semirings, the situation is completely different. The free continuous commutative idempotent semirings are the semirings of commutative languages, i.e., the semirings $L_{\Sigma^\oplus} = P(\Sigma^\oplus)$, where $\Sigma^\oplus$ is the free commutative monoid of commutative words generated by Σ. (When Σ is finite and has k elements, $L_{\Sigma^\oplus}$ is isomorphic to $P(\mathbb{N}^k)$.) By Parikh's theorem [25], the context-free sets included in $\Sigma^\oplus$ are the same as the rational sets, or the semilinear sets, and equality of semilinear sets is decidable and is logspace complete for Π_2^p, the second level of the polynomial time hierarchy [15]. It follows that the equivalence problem of commutative context-free grammars is decidable, see also [10]. In fact, as shown in [16], the equivalence problem for both commutative context-free grammars and commutative rational expressions are solvable in nondeterministic exponential time. Thus, the equational theory of continuous idempotent commutative semirings is also decidable in nondeterministic exponential time. An infinite basis of the rational identities for commutative languages was given by Redko [28] (see also the treatments by Salomaa [29] and Conway [6]). The language of rational terms is a sublanguage of the language of μ-terms. As a corollary of our equational proof of Parikh's theorem, and of Redko's axiomatization, we give a basis of identities of commutative languages and thus of continuous commutative idempotent semirings for the *full* language of μ-terms.

References

[1] H. Bekić. Definable operations in general algebras, and the theory of automata and flowcharts. *Technical Report*, IBM Laboratory, Vienna, 1969.

[2] S. L. Bloom and Z. Ésik. Floyd-Hoare logic in iteration theories. *J. Assoc. Comput. Mach.*, 38(1991), 887–934.

[3] S. L. Bloom and Z. Ésik. Program correctness and matricial iteration theories. In: *Proc. Mathematical Foundations of Programming Semantics'91*, LNCS 598, Springer-Verlag, 1992, 457–475.

[4] S. L. Bloom and Z. Ésik. *Iteration Theories.* Springer-Verlag, 1993.

[5] S. Bozapalidis. Equational elements in additive algebras. *Theory Comput. Syst.*, 32(1999), 1–33.

[6] J. Conway. *Regular Algebra and Finite Machines.* Chapman and Hall, 1971.

[7] J. W. De Bakker and D. Scott, A theory of programs, *IBM Seminar*, Vienna, 1969.

[8] Z. Ésik. Group axioms for iteration. *Inform. and Comput.*, 148(1999), 131–180.

[9] Z. Ésik and H. Leiss. In preparation.

[10] S. Ginsburg. *The Mathematical Theory of Context-Free Languages.* McGraw-Hill, 1966.

[11] A. Ginzburg. *Algebraic Theory of Automata.* Academic Press, New York-London, 1968.

[12] J. S. Golan. *Semirings and their Applications.* Kluwer Academic Publishers, Dordrecht, 1999.

[13] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, Reading, Mass., 1979.

[14] M. W. Hopkins and D. Kozen. Parikh's theorem in commutative Kleene algebra. In: *Proc. IEEE Conf. Logic in Computer Science (LICS'99)*, IEEE Press, 1999, 394–401.

[15] D. T. Huynh. The complexity of semilinear sets. *Elektron. Informationsverarb. Kybernet*, 18(1982), 291–338.

[16] D. T. Huynh. The complexity of equivalence problems for commutative grammars. *Inform. and Control*, 66(1985), 103–121.

[17] D. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. *Inform. and Comput.*, 110(1994), 366–390.

[18] D. Kozen. On Hoare logic and Kleene algebra with tests. In: *Proc. IEEE Conf. Logic in Computer Science (LICS'99), IEEE, July 1999, 167–172.*

[19] D. Krob. Complete systems of B-rational identities. *Theoret. Comput. Sci.*, 89(1991), 207–343.

[20] W. Kuich. The Kleene and the Parikh theorem in complete semirings. In: *Proc. ICALP '87*, LNCS 267, Springer-Verlag, 212–225.

[21] W. Kuich. Gaussian elimination and a characterization of algebraic power series. In: *Proc. Mathematical Foundations of Computer Science, 1998*, LNCS 1450, Springer, Berlin, 1998, 512–521.

[22] W. Kuich and A. Salomaa. *Semirings, Automata, Languages.* Springer-Verlag, Berlin, 1986.

[23] E. G. Manes and M. A. Arbib. *Algebraic Approaches to Program Semantics.* Springer-Verlag, New York, 1986.

[24] D. Niwiński. On fixed-point clones (extended abstract). In: *Automata, Languages and Programming, Rennes, 1986*, LNCS 226, Springer, 1986. 464–473.

[25] R. J. Parikh. On context-free languages. *J. Assoc. Comput. Mach.*, 13(1966), 570–581.

[26] D. M. R. Park. Fixed point induction and proofs of program properties. In: *Machine Intelligence* 5, Edinburgh Univ. Press, 1970, 59–78.

[27] D. L. Pilling. Commutative regular equations and Parikh's theorem. *J. London Math. Soc.*, 6(1973), 663–666.

[28] V. N. Redko. On the algebra of commutative events. (Russian) *Ukrain. Mat. Ž.* 16(1964), 185–195.

[29] A. Salomaa. *Theory of Automata.* Pergamon Press, 1969.

[30] I. Takanami and N. Honda. A characterization of Parikh's theorem and semilinear sets by commutative semigroups with length. *Electronics and Communications in Japan*, 52(1969), 179–184.