

Combinatorial algebra: syntax and semantics, by Mark V. Sapir, with contributions by Victor S. Guba and Mikhail V. Volkov, Springer Monographs in Mathematics, Springer, Cham, 2014, xvi+355 pp., ISBN 978-3-319-08030-7 and 978-3-319-08031-4

Algebra—*الحبر والمقابلة*, the “tearing-apart and reassembling”—is often merely seen as the art of stripping meaning away from mathematical facts, so as to arrive at more essential properties that apply more widely. In this manner, one establishes links between seemingly unrelated domains.

As a simple example, one easily shows following Lagrange that, in any finite group, the order of every element divides the order of the group. Applied to the non-zero residue classes modulo a prime p , one deduces that $a^{p-1} \equiv 1 \pmod{p}$ for every a coprime to p .

This is the *syntactic* approach to algebra. Another important trend, the “re-assembling”, uses mathematical objects from diverse origins to solve problems in a specific area.

Returning to the previous example, every group acts by right-multiplication on itself. This action is *regular*, so every element acts as a product of cycles all of the same length, and Lagrange’s statement follows. The *semantic* approach encouraged us to look at meaningful realisations of our group, in this case as a group of permutations.

The book under review by Mark Sapir is an enticing collection of statements highlighting the interplay between the syntactic and semantic aspects of algebra. Its parts are (almost) titled “words”, “semigroups”, “rings”, “groups”; one feels that the author suffered very much in being forced to write the text linearly, so rich are the interconnections between these topics.

There is another opposition, which somewhat follows the syntax/semantics dichotomy: mathematical objects may be explored via their quotients and via their subobjects. Given an object X , to check that Y is a quotient of X , it is usually preferable to have a syntactic description of X ; for groups or algebras, one would want a presentation of X by generators and relations, the archetypal syntactic description of X . To check that Y is a subobject of X , one would rather want a concrete description of X ; for groups or algebras, via their action on some well-understood object.

Thus free groups are best viewed syntactically as sets F_A of reduced words over $A \sqcup A^{-1}$ with concatenation and reduction as operation and semantically as free actions on trees.

1. THE COMBINATORIAL ACTORS

Rather than presenting the main algebraic actors—semigroups, algebras, groups as in the book under review—I will present them together, so as to better point out their connections.

Mathematical objects are constructed out of elementary parts. Words and graphs stand out among these. Fix a finite set A called the *alphabet*; one may then consider

2010 *Mathematics Subject Classification*. Primary 05E15.

the set A^* of finite sequences of elements of A , called *words*, and subsets of A^* , called *languages*. There are then $A^{\mathbb{N}}, A^{-\mathbb{N}}, A^{\mathbb{Z}}$ the spaces of right-infinite, left-infinite, and bi-infinite words. These infinite sequences determine a *language of subwords*.

The latter space $A^{\mathbb{Z}}$ is particularly interesting because it is a compact set equipped with the dynamics of an invertible transformation T acting by shift. The closed T -invariant subspaces of $A^{\mathbb{Z}}$, and in particular the minimal ones, received particular attention and are called *subshifts*. A subshift $X \subseteq A^{\mathbb{Z}}$ may be studied via its associated language of *forbidden words*: these are the words $w \in A^*$ that do not occur as a subword of any element of X . Conversely, every language $L \subset A^*$ determines a subshift: the set of bi-infinite words none of whose subwords belongs to L . Without loss of generality, the language L is an *ideal*, $L = A^*L_0A^*$ for a *generating* set of minimal words $L_0 \subset L$. For example, the Fibonacci shift $\Phi \subset \{0, 1\}^{\mathbb{Z}}$ consists of all bi-infinite words without two consecutive 1's; its forbidden language is generated by $\{11\}$.

A single bi-infinite word may also be used to produce a subshift, as the closure of its set of T -translates. Thus for example iterating the substitution $0 \mapsto 01, 1 \mapsto 0$ infinitely many times on 0.1 (the “.” marks the position of the origin) yields an infinite word in $\{0, 1\}^{\mathbb{Z}}$, and the closure of its T -orbit is a subshift Φ_0 of the Fibonacci shift Φ , with more forbidden words ($11, 000, \dots$, and also all cubes).

A *directed graph* is specified by a vertex set V , an edge set E , and two maps $E \rightarrow V$ called *head* and *tail*. A path in a graph is a sequence of edges whose heads and tails match. The set of bi-infinite walks in a graph naturally defines a subshift $X \subseteq E^{\mathbb{Z}}$, called its *edge subshift*. If the graph is finite, the forbidden language of X is finite and X is called a *subshift of finite type*.

More generally, a graph's edges may be labeled by a finite set A , and it is then called an *automaton*. The set of labels read along bi-infinite paths in an automaton defines a subshift of $A^{\mathbb{Z}}$; it is also defined by the forbidden language of all illegal transitions in the automaton. If an initial vertex is fixed in the automaton, the set of words read along paths starting at the initial vertex defines a *regular* language, the language *accepted* by the automaton. One gains flexibility, but no generality, by restricting to paths with given initial and given final vertices. For example, the Fibonacci shift Φ is encoded by the following automaton.



2. THE ALGEBRAIC ACTORS

First and foremost, the set of words A^* is naturally a semigroup, under concatenation.

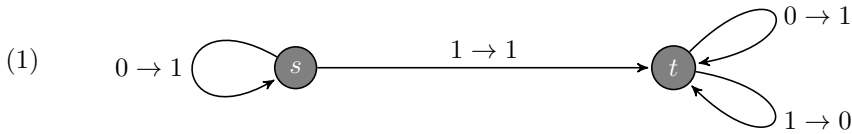
From a set of forbidden words, one may construct a *semigroup with 0*: the quotient of the free semigroup-with-zero $A^* \cup \{0\}$ by the relation “ $w = 0$ ” for every forbidden word. Conversely, given S a semigroup with generating set A , one constructs the *Cayley graph* of S as the oriented graph with vertex set S and edge set $S \times A$, where the edge (s, a) goes from s to sa . Every element of S may be written as a word in A^* , and an expression of minimal length is called a *geodesic*.

Bi-infinite geodesics are maps $A^{\mathbb{Z}} \rightarrow S$ all of whose subwords are geodesic and naturally define a subshift.

One may directly start with S a semigroup with 0 and a generating set A of S , and define an associated subshift $X \subseteq A^{\mathbb{Z}}$ as the set of all bi-infinite words over A all of whose subwords multiply to a non-trivial element of S . For example, starting with $S = \langle a, b \mid b^2 = 0 \rangle_+$, one obtains again the Fibonacci subshift Φ .

From a graph $G = (V, E)$, one constructs a semigroup-with-zero $A(G) = V \cup \{0\}$ by declaring $v \cdot w = v$ if there is an edge from v to w , and $v \cdot w = 0$ otherwise.

Automata may also be studied algebraically. An automaton is *deterministic* if at every vertex the collection of labels on its outgoing edges is in bijection with A . The set of self-maps of V defined by the maps “follow label a ” for all $a \in A$ generates a semigroup called the *syntactic semigroup* of the automaton. A deterministic automaton with a second label in A called the *output* is called a *Mealy transducer*; endowed with an initial vertex, it produces a transformation of A^* by the rule “given a word $w \in A^*$, follow the path labeled w in the automaton, and map w to the output word read along the same path.” Quite complicated semigroups may be produced by very small automata, such as the following automaton.



Here the transformation t exchanges 0 and 1 in every word, and the transformation s sends, e.g., 001100 to 111011. The semigroup generated by $\{s, t\}$ has “intermediate word growth”; see below.

Let \mathbb{k} be a field. From a semigroup S one constructs the *semigroup algebra* $\mathbb{k}S$, the set of \mathbb{k} -linear combinations of elements of S , with natural addition and multiplication. Conversely, multiplicatively closed subsets of associative algebras give semigroups. The semigroup algebra $\mathbb{k}A^*$ is the algebra of non-commutative polynomials in the variables A often written $\mathbb{k}\langle A \rangle$. Its completion $\mathbb{k}\langle\langle A \rangle\rangle = \mathbb{k}A^*$ is the algebra of non-commutative power series with variables in A .

3. GOALS

Some fundamental questions have dominated the field of algebra, and have served as useful lighthouses in directing research towards fertile grounds.

One of them is the study of laws and identities in algebraic objects. The proper setting is that of *universal algebra*, in which objects, called *algebras*, are endowed with a fixed collection of operations of various arities.

One probes an algebra X by considering all equalities of the form $t = t'$ with t, t' in the free algebra (free semigroup A^* , free associative algebra $\mathbb{k}\langle A \rangle$, free group F_A) that hold in X , namely that become equalities in X under arbitrary substitutions of elements of X for variables in A . For example, X is commutative if $xy = yx$ holds universally in X . Conversely, one may start with a collection Σ of identities, and study the *variety* of algebras that satisfy Σ .

Garrett Birkhoff's fundamental result defines varieties without reference to identities: a class of algebras is a variety if and only if it is closed under taking subalgebras, homomorphic images, and arbitrary Cartesian products. Thus a variety may be *generated* by a collection of algebras and *defined* by a collection of identities.

A variety is *finitely based* if it may be defined by a finite set of identities. The variety generated by a finite group, or a finite associative ring, is finitely based. A fundamental question is, Which varieties are finitely based?

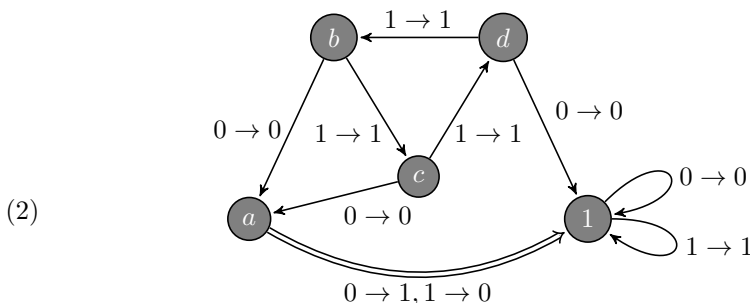
One of the most fruitful questions asked about varieties is, *When are they locally finite?* (In other words, When is every finitely generated algebra in the variety finite?) Burnside's question asks, for fixed n , whether every finitely generated group in which every element has order dividing n is itself finite; or, in other words, whether the variety defined by $x^n = 1$ is locally finite. This has been answered in the negative for n large enough, and in the positive for $n = 4$ and $n = 6$, but is still open for $n = 5$.

A locally finite variety is *inherently non-finitely based* if for every finite subset of its identities there is an infinite finitely generated algebra satisfying these identities. A result by Baker, McNulty, and Werner characterizes those inherently non-finitely based varieties generated by a semigroup of the form $A(G)$ in terms of four induced subgraphs that G may contain. The proof is a beautiful study of the closed orbits of the dynamical system $B^{\mathbb{Z}}$ for a finite algebra B .

Moving beyond varieties, one may call an algebra *algebraic* if every cyclic (singly generated) subalgebra is finite. A group is usually called *torsion* in that case, and the generalized Burnside question asks whether there are infinite, finitely generated torsion groups. This has been answered positively, even among residually finite groups (namely, groups admitting enough finite quotients to distinguish any two elements). Remarkably, there are both syntactic and semantic proofs of this result; see below. In universal algebra with 0 (associative algebras, semigroups-with-zero) one calls an algebra *nil* if every element $x \in A$ satisfies $x^n = 0$ for some n .

Golod gave a syntactic proof of the existence of finitely generated, infinite-dimensional nil associative algebras: he constructed them as quotients of the free associative algebra $\mathbb{F}_p\langle\langle A \rangle\rangle$ by relations of the form $w^{n(w)} = 0$ for all $w \in A^*$, with $n(w) \in \mathbb{N}$ large enough so that the quotient is infinite dimensional. The sub-semigroup generated by $\{1 + a : a \in A\}$ is a finitely generated, infinite torsion group.

There is no semantic construction of finitely generated, infinite-dimensional nil algebras, but there is such a construction of finitely generated, infinite torsion groups, due to Aleshin and simplified by Grigorchuk. Grigorchuk's example is a group of permutations of A^* generated by the following remarkably small Mealy transducer:



For a finitely generated algebra (group, semigroup, associative algebra, ...), one may consider its *growth*, namely the function counting the number of (*ad lib.* linearly independent) elements that can be produced as products of at most n generators. It is natural to consider this growth function up to rescaling on its argument, so as to remove the dependency on the choice of generating set; thus all exponential functions are treated as equivalent, and polynomially growing functions are equivalent to their leading monomial. Fundamental questions are then, Which kinds of growth functions may occur? and Are there gaps in the spectrum of growth functions? For example, there is no associative algebra with growth strictly between constant and linear, nor between linear and quadratic; and there is no group with growth strictly between polynomial and $n^{(\log n)^{1/100}}$; conjecturally, the gap extends to $\exp(\sqrt{n})$. On the other hand, there is a group (a small variant of Grigorchuk's example from (2)) with growth $\exp(n^{0.7675\dots})$, and every sufficiently smooth function between $\exp(n^{0.7675\dots})$ and $\exp(n)$ is asymptotic to the growth function of a group.

There is an even wider spectrum of growth of semigroups; many superquadratic functions, and every sufficiently smooth function between $n^{\log n}$ and $\exp(n)$ is asymptotic to the growth function of a semigroup. For example, one may (syntactically) consider the monoid with presentation $\langle y, z \mid z^{p+1}y^{p+1}z = z^{p+1}y^p \text{ for all } p \geq 0 \rangle$, and note that its elements may be written uniquely in the *normal form* $y^{a_1}z^{a_2}\dots y^{a_n}$ for a partition (a_1, \dots, a_n) ; the Hardy–Ramanujan estimates on the growth of partitions directly give estimates on the growth of the monoid, of the form $\exp(\sqrt{n})$. Semantically, one may realize this monoid as a set of transformations of A^* generated by a small Mealy transducer very close to (1), or as a set of matrices with integer coefficients.

More generally, subsets of algebras may also be assigned a growth function called their *relative growth*; to compute the growth of a semigroup S with generating set A , it may be good to find a *geodesic normal form* for S : for every $s \in S$ a choice of a minimal-length expression in A^* representing it; the set L of such normal forms is a language $L \subseteq A^*$, whose growth is that of S . In turn, this normal form may be obtained, in favourable cases, by applying *rewriting rules*, namely substitutions of the form $u \rightarrow v$ on subwords of A^* for some $u, v \in A^*$ that define the same element of S . Some semigroups (and more generally algebras) admit a finite set of rewriting rules whose iterated application eventually leads to a normal form.

Subshifts are naturally measured by their *complexity*, which is simply the relative growth of the language of subwords of its elements. This is also the growth of the semigroup naturally associated with the subshift. For example, the complexity of the Fibonacci subshift Φ is the Fibonacci sequence; while that of its subshift Φ_0 is linear (it is a “Sturmian subshift”).

Yet another kind of growth function associated with an algebra measures the growth of its relations: for a semigroup S generated by A , consider the natural map $\pi: A^* \twoheadrightarrow S$ and the relative growth of the language $\{(u, v) \in (A \times A)^* \mid \pi(u) = \pi(v)\}$ of its relations; for a group $G = \langle A \rangle$, consider the relative growth of $\pi^{-1}(1)$ with $\pi: F_A \twoheadrightarrow G$; for an associative algebra R generated by a subset A , consider the relative growth of $\pi^{-1}(0)$ with $\pi: \mathbb{k}\langle A \rangle \twoheadrightarrow R$. An important property, *amenability* of groups, may be defined combinatorially as the condition that $\pi^{-1}(1)$ has exponentially small growth relative to F_A , or equivalently relative to A^* .

4. HIGHLIGHTS OF THE BOOK

It is time to turn to more specifics about what the reader may find in the book under review. A large number of important results are stated and, more importantly, proven. Not only are the proof techniques interesting in themselves, they help very much in bridging the various topics under consideration.

The book starts by word avoidance problems, such as the classical Thue–Morse cube-free sequence, and introduces Zimin words and their use in characterizing repetition-free languages. These Zimin words reappear throughout the text as a *leitmotiv*.

The *road colouring problem* asks whether every aperiodic graph with constant out-degree may be labeled so as to give a *reset* automaton—an automaton that admits a word leading to a given vertex, regardless of the starting vertex, or equivalently whose syntactic monoid admits a right 0. This was proven recently by Trahtman, and his (combinatorial) proof is included with the original motivation towards algorithmically deciding isomorphism of subshifts of finite type.

An important, but still hard, result in group theory is the construction, by Adian, of infinite finitely generated groups of finite exponent (“Burnside groups”). A complete proof would go beyond the scope of the book, but a useful roadmap is given, highlighting the main features of the argument.

Within the topic of associative algebras, the author gives a readable account of Shirshov’s height theorem, and as its consequence Kaplansky’s result that algebraic algebras with identity are finite dimensional.

By necessity, the author had to select which results to include. One gets a feeling of his favourites, including “diagram groups”, which are groups whose elements are reduced diagrams expressing equalities derived from a semigroup presentation. This class contains classical examples such as free, and free abelian groups as well as more exotic ones such as Thompson’s group F : the group of piecewise-linear bijections of the interval $[0, 1]$ with slopes a power of 2 and dyadic breakpoints. Amenability of F is a well-known open question.

Many of the proofs given throughout the text are quite concentrated and “to the point”; verifications are left to the reader, with copious use of “(prove it!)” and “(check!)” interjections. The large number of typographical mistakes provide a pleasant extra exercise set to the reader.

LAURENT BARTHOLDI

ÉCOLE NORMALE SUPÉRIEURE, PARIS, FRANCE

E-mail address: laurent.bartholdi@ens.fr