Equational theories

Terence Tao and the other contributors of the Equational Theory Project

October 2, 2024

Chapter 1

Basic theory of magmas

Definition 1.1 (Magma)**.** A *magma* is a set G equipped with a binary operation $\circ : G \times G \to G$. A *homomorphism* $\varphi : G \to H$ between two magmas is a map such that $\varphi(x \circ y) = \varphi(x) \circ \varphi(y)$ for all $x, y \in G$. An *isomorphism* is an invertible homomorphism.

Groups, semi-groups, and monoids are familiar examples of magmas. However, in general we do not expect magmas to have any associative properties.

A magma is called *empty* if it has cardinality zero, *singleton* if it has cardinality one, and *non-trivial* otherwise.

The number of magma structures on a set G of cardinality n is of course n^{n^2}, which is [1]

$$1, 1, 16, 19683, 4294967296, 298023223876953125, \ldots$$

(OEIS A002489). Up to isomorphism, the number of finite magmas of cardinality n up to isomorphism is the slightly slower growing sequence

$$1, 1, 10, 3330, 178981952, 2483527537094825, 14325590003318891522275680, \ldots$$

(OEIS A001329).

Definition 1.2 (Free Magma)**.** The *free magma* M_X generated by a set X (which we call an *alphabet*) is the set of all finite formal expressions built from elements of X and the operation $\circ$. An element of M_X will be called a *word* with alphabet X. The *order* of a word is the number of $\circ$ symbols needed to generate the word. Thus for instance X is precisely the set of words of order 0 in M_X.

For sake of concreteness, we will take the alphabet X to default to the natural numbers $\mathbb{N}$ if not otherwise specified.

For instance, if $X = \{0, 1\}$, then M_X would consist of the following words:

- 0, 1 (the words of order 0);
- $0 \circ 0$, $0 \circ 1$, $1 \circ 0$, $1 \circ 1$ (the words of order 1);
- $0 \circ (0 \circ 0)$, $0 \circ (0 \circ 1)$, $0 \circ (1 \circ 0)$, $0 \circ (1 \circ 1)$, $1 \circ (0 \circ 0)$, $1 \circ (0 \circ 1)$, $1 \circ (1 \circ 0)$, $1 \circ (1 \circ 1)$, $(0 \circ 0) \circ 0$, $(0 \circ 0) \circ 1$, $(0 \circ 1) \circ 0$, $(0 \circ 1) \circ 1$, $(1 \circ 0) \circ 0$, $(1 \circ 0) \circ 1$, $(1 \circ 1) \circ 0$, $(1 \circ 1) \circ 1$ (the words of order 2);

[1] All sequences start from $n = 0$ unless otherwise specified.

- etc.

Lemma 1.3. *For a finite alphabet X, the number of words of order n is $C_n|X|^{n+1}$, where C_n is the n^{th} Catalan number and X is the cardinality of X.*

Proof. Follows from standard properties of Catalan numbers. $\square$

The first few Catalan numbers are

$$1, 1, 2, 5, 14, 42, 132, \ldots$$

(OEIS A000108).

Definition 1.4 (Induced homomorphism)**.** Given a function $f : X \to G$ from an alphabet X to a magma G, the *induced homomorphism* $\varphi_f : M_X \to G$ is the unique extension of f to a magma homomorphism. Similarly, if $\pi\colon X \to Y$ is a function, we write $\pi_*\colon M_X \to M_Y$ for the unique extension of π to a magma homomorphism.

For instance, if $f : \{0,1\} \to G$ maps $0, 1$ to x, y respectively, then

$$\varphi_f(0 \circ 1) = x \circ y$$

$$\varphi_f(1 \circ (0 \circ 1)) = y \circ (x \circ y)$$

and so forth. If $\pi\colon \mathbb{N} \to \mathbb{N}$ is the map $\pi(n) := n + 1$, then

$$\pi_*(0 \circ 1) = 1 \circ 2$$

$$\pi_*(1 \circ (0 \circ 1)) = 2 \circ (1 \circ 2)$$

and so forth.

Definition 1.5 (Law)**.** Let X be a set. A *law* with alphabet X is a formal expression of the form $w \simeq w'$, where $w, w' \in M_X$ are words with alphabet X (thus one can identify laws with alphabet X with elements of $M_X \times M_X$). A magma G *satisfies* the law $w \simeq w'$ if we have $\varphi_f(w) = \varphi_f(w')$ for all $f : X \to G$, in which case we write $G \models w \simeq w'$.

Thus, for instance, the commutative law

$$0 \circ 1 \simeq 1 \circ 0 \tag{1.1}$$

is satisfied by a magma G if and only if

$$x \circ y = y \circ x \tag{1.2}$$

for all $x, y \in G$. We refer to (1.2) as the *equation* associated to the law (1.1). One can think of equations as the “semantic” intrepretation of a “syntactic” law. However, we shall often abuse notation and a law with its associated equation thus we shall (somewhat carelessly) also refer to (1.2) as “the commutative law” (rather than “the commutative equation”).

Lemma 1.6 (Pushforward)**.** *Let $w \simeq w'$ be a law with some alphabet X, G be a magma, and $\pi : X \to Y$ be a function. If $G \models w \simeq w'$, then $G \models \pi_*(w) \simeq \pi_*(w')$. In particular, if π is a bijection, the statements If $G \models w \simeq w'$, then $G \models \pi_*(w) \simeq \pi_*(w')$ are equivalent.*

If π is a bijection, we will call $\pi_*(w) \simeq \pi_*(w')$ a *relabeling* of the law $w \simeq w'$. Thus for instance

$$5 \circ 7 \simeq 7 \circ 5$$

is a relabeling of the commutative law (1.1). By the above lemma, relabeling does not affect whether a given magna satisfies a given law.

Proof. Trivial. □

Lemma 1.7 (Equivalence)**.** *Let G be a magma and X be an alphabet. Then the relation $G \models w \simeq w'$ is an equivalence relation on M_X.*

Proof. Trivial. □

Define the total order of a law $w \simeq w'$ to be the sum of the orders of w and w'.

Lemma 1.8 (Counting laws up to relabeling)**.** *Up to relabeling, the number of laws $w \simeq w'$ of total order n is $C_{n+1}B_{n+2}$.*

Proof. Follows from the properties of Catalan and Bell numbers. □

The first few Bell numbers (starting from $n = 0$) are

$$1, 1, 2, 5, 15, 52, 203, \dots$$

(OEIS A000110).

The sequence in Lemma 1.8 is

$$2, 10, 75, 728, 8526, 115764, \dots$$

(OEIS A289679).

Now we would also like to count laws up to relabeling and symmetry.

Lemma 1.9 (Counting laws up to relabeling and symmetry)**.** *Up to relabeling and symmetry, the number of laws $w \simeq w'$ of total order n is*

$$C_{n+1}B_{n+2}/2$$

when n is odd, and

$$(C_{n+1}B_{n+2} + C_{n/2}(2D_{n+2} - B_{n+2}))/2$$

when n is even, where D_n is the number of partitions of $[n]$ up to reflection.

Proof. Elementary counting. □

The sequence D_n is (starting from $n = 0$)

$$1, 1, 2, 4, 11, 32, 117, \dots$$

(OEIS A103293), and the sequence in Lemma 1.9 is (starting from $n = 0$)

$$2, 5, 41, 364, 4294, 57882, 888440, \dots$$

We can also identify all laws of the form $w \simeq w$ with the trivial law $0 \simeq 0$. The number of such laws of total order n is zero if n is odd, and $C_{n/2}B_{n/2+1}$ if n is even. We conclude:

Lemma 1.10 (Counting laws up to relabeling, symmetry, and triviality)**.** *Up to relabeling, symmetry, and triviality, the number of laws of total order n is*

$$C_{n+1}B_{n+2}/2$$

if n is odd, 2 if $n = 0$, and

$$(C_{n+1}B_{n+2} + C_{n/2}(2D_{n+2} - B_{n+2}))/2 - C_{n/2}B_{n/2+1}$$

if $n \geq 2$ is even.

Proof. Routine counting. $\square$

This sequence is

$$2, 5, 39, 364, 4284, 57882, 888365, \ldots$$

In particular, up to relabeling, symmetry, and triviality, there are exactly 4694 laws of total order at most 4. A list can be found here. A script for generating them may be found here. The list is sorted by the total number of operations, then by the number of operations on the LHS. Within each such class we define an order on expressions by variable < operation, and lexical order on variables.

Chapter 2

Subgraph laws

In this project we study the 4694 laws (up to symmetry and relabeling) of total order at most 4.

Selected laws of interest are listed below, as well as in this file. Laws in this list will be referred to as "subgraph equations", as we shall inspect the subgraph of the implication subgraph induced by these equations.

Definition 2.1 (Equation 1)**.** Equation 1 is the law $0 \simeq 0$ (or the equation $x = x$).

This is the trivial law, satisfied by all magmas. It is self-dual.

Definition 2.2 (Equation 2)**.** Equation 2 is the law $0 \simeq 1$ (or the equation $x = y$).

This is the singleton law, satisfied only by the empty and singleton magmas. It is self-dual.

Definition 2.3 (Equation 3)**.** Equation 3 is the law $0 \simeq 0 \circ 0$ (or the equation $x = x \circ x$).

This is the idempotence law. It is self-dual.

Definition 2.4 (Equation 4)**.** Equation 4 is the law $0 \simeq 0 \circ 1$ (or the equation $x = x \circ y$).

This is the left absorption law.

Definition 2.5 (Equation 5)**.** Equation 5 is the law $0 \simeq 1 \circ 0$ (or the equation $x = y \circ x$).

This is the right absorption law (the dual of Definition 2.4).

Definition 2.6 (Equation 6)**.** Equation 6 is the law $0 \simeq 1 \circ 1$ (or the equation $x = y \circ y$).

This law is equivalent to the singleton law.

Definition 2.7 (Equation 7)**.** Equation 7 is the law $0 \simeq 1 \circ 2$ (or the equation $x = y \circ z$).

This law is equivalent to the singleton law.

Definition 2.8 (Equation 8)**.** Equation 8 is the law $0 \simeq 0 \circ (0 \circ 0)$ (or the equation $x = x \circ (x \circ x)$).

Definition 2.9 (Equation 14)**.** Equation 14 is the law $0 \simeq 1 \circ (0 \circ 1)$ (or the equation $x = y \circ (x \circ y)$).

Appears in Problem A1 from Putnam 2001.

Definition 2.10 (Equation 23)**.** Equation 23 is the law $0 \simeq (0 \circ 0) \circ 0$ (or the equation $x = (x \circ x) \circ x$).

This is the dual of Definition 2.8.

Definition 2.11 (Equation 29)**.** Equation 29 is the law $0 \simeq (1 \circ 0) \circ 1$ (or the equation $x = (y \circ x) \circ y$).

Appears in Problem A1 from Putnam 2001. Dual to Definition 2.9.

Definition 2.12 (Equation 38)**.** Equation 38 is the law $0 \circ 0 \simeq 0 \circ 1$ (or the equation $x \circ x = x \circ y$).

This law asserts that the magma operation is independent of the second argument.

Definition 2.13 (Equation 39)**.** Equation 39 is the law $0 \circ 0 \simeq 1 \circ 0$ (or the equation $x \circ x = y \circ x$).

This law asserts that the magma operation is independent of the first argument (the dual of Definition 2.12).

Definition 2.14 (Equation 40)**.** Equation 40 is the law $0 \circ 0 \simeq 1 \circ 1$ (or the equation $x \circ x = y \circ y$).

This law asserts that all squares are constant. It is self-dual.

Definition 2.15 (Equation 41)**.** Equation 41 is the law $0 \circ 0 \simeq 1 \circ 2$ (or the equation $x \circ x = y \circ z$).

This law is equivalent to the constant law, Definition 2.19.

Definition 2.16 (Equation 42)**.** Equation 42 is the law $0 \circ 1 \simeq 0 \circ 2$ (or the equation $x \circ y = x \circ z$).

Equivalent to Definition 2.12.

Definition 2.17 (Equation 43)**.** Equation 43 is the law $0 \circ 1 \simeq 1 \circ 0$ (or the equation $x \circ y = y \circ x$).

The commutative law. It is self-dual.

Definition 2.18 (Equation 45)**.** Equation 45 is the law $0 \circ 1 \simeq 2 \circ 1$ (or the equation $x \circ y = z \circ y$).

This is the dual of Definition 2.16.

Definition 2.19 (Equation 46)**.** Equation 46 is the law $0 \circ 1 \simeq 2 \circ 3$ (or the equation $x \circ y = z \circ w$).

The constant law: all products are constant. It is self-dual.

Definition 2.20 (Equation 168)**.** Equation 168 is the law $0 \simeq (1 \circ 0) \circ (0 \circ 2)$ (or the equation $x = (y \circ x) \circ (x \circ z)$).

The law of a central groupoid. It is self-dual.

Definition 2.21 (Equation 381)**.** Equation 381 is the law $0 \circ 1 \simeq (0 \circ 2) \circ 1$ (or the equation $x \circ y = (x \circ z) \circ y$).

Appears in Putnam 1978, Problem A4, part (b).

Definition 2.22 (Equation 387)**.** Equation 387 is the law $0 \circ 1 \simeq (1 \circ 1) \circ 0$ (or the equation $x \circ y = (y \circ y) \circ x$).

Definition 2.23 (Equation 3722)**.** Equation 3722 is the law $0 \circ 1 \simeq (0 \circ 1) \circ (0 \circ 1)$ (or the equation $x \circ y = (x \circ y) \circ (x \circ y)$).

Appears in Putnam 1978, Problem A4, part (a). It is self-dual.

Definition 2.24 (Equation 3744)**.** Equation 3744 is the law $0 \circ 1 \simeq (0 \circ 2) \circ (3 \circ 1)$ (or the equation $x \circ y = (x \circ z) \circ (w \circ y)$).

This law is called a "bypass operation" in Putnam 1978, Problem A4. It is self-dual.

Definition 2.25 (Equation 4512)**.** Equation 4512 is the law $0 \circ (1 \circ 2) \simeq (0 \circ 1) \circ 2$ (or the equation $x \circ (y \circ z) = (x \circ y) \circ z$).

The associative law. It is self-dual.

Definition 2.26 (Equation 4513)**.** Equation 4513 is the law $0 \circ (1 \circ 2) \simeq (0 \circ 1) \circ 3$ (or the equation $x \circ (y \circ z) = (x \circ y) \circ w$).

Definition 2.27 (Equation 4522)**.** Equation 4522 is the law $0 \circ (1 \circ 2) \simeq (0 \circ 3) \circ 4$ (or the equation $x \circ (y \circ z) = (x \circ w) \circ u$).

Dual to Definition 2.29.

Definition 2.28 (Equation 4564)**.** Equation 4564 is the law $0 \circ (1 \circ 2) \simeq (3 \circ 1) \circ 2$ (or the equation $x \circ (y \circ z) = (w \circ y) \circ z$).

Dual to Definition 2.26.

Definition 2.29 (Equation 4579)**.** Equation 4579 is the law $0 \circ (1 \circ 2) \simeq (3 \circ 4) \circ 2$ (or the equation $x \circ (y \circ z) = (w \circ u) \circ z$).

Dual to Definition 2.27.

Definition 2.30 (Equation 4582)**.** Equation 4582 is the law $0 \circ (1 \circ 2) \simeq (3 \circ 4) \circ 5$ (or the equation $x \circ (y \circ z) = (w \circ u) \circ v$).

This law asserts that all triple constants (regardless of bracketing) are constant.
Here is a more complicated law, introduced by Kisielewicz [**?**]:

Definition 2.31 (Equation 374794)**.** Equation 374794 is the law $0 \simeq (((1 \circ 1) \circ 1) \circ 0) \circ ((1 \circ 1) \circ 2)$ (or the equation $x = (((y \circ y) \circ y) \circ x) \circ ((y \circ y) \circ z)$).

We will be interested in seeing which laws imply which other laws, in the sense that magmas obeying the former law automatically obey the latter. We will also be interested in *anti-implications* showing that one law does *not* imply another, by producing examples of magmas that obey the former law but not the latter.

The singleton or empty magma obeys all equational laws. One can ask whether an equational law admits nontrivial finite or infinite models. An *Austin law* is a law which admits infinite models, but no nontrivial finite models. Austin [**?**] established the first such law, namely the order 9 law

$$(((1 \circ 1) \circ 1) \circ 0) \circ (((1 \circ 1) \circ ((1 \circ 1) \circ 1)) \circ 2) = 0.$$

A shorter Austin law (order 6) was established in [**?**]:

Theorem 2.32 (Kisielewicz theorem)**.** *Definition 2.31 is an Austin law.*

Proof. Suppose for contradiction that we have a non-trivial model of Definition 2.31. Write $y^2 := y \circ y$ and $y^3 := y^2 \circ y$. For any y, z, introduce the functions $f_y : x \mapsto y^3 \circ x$ and $g_{yz} : x \mapsto x \circ (y^2 \circ z)$. Definition 2.31 says that g_{yz} is a left-inverse of f_y, hence by finiteness these are inverses and g_{yz} is independent of z. In particular

$$f(y^3) = g_{yy}(y^3) = g_{yz}(y^3) = f(y^2 \circ z)$$

and hence $y^2 \circ z$ is independent of z. Thus

$$f_y(x) = (y^2 \circ y) \circ x = (y^2 \circ y^2) \circ x$$

is independent of x. As f_y is invertible, this forces the magma to be trivial, a contradiction.

To construct an infinite magma, take the positive integers $\mathbb{Z}^+$ with the operation $x \circ y$ defined as

- 2^x if $y = x$;
- 3^y if $x = 1 \neq y$;
- $\min(j, 1)$ if $x = 3^j$ and $y \neq x$; and
- 1 otherwise.

Then $y^2 = 2^y$, $y^3 = 1$, and $y^2 \circ z$ a power of two for all y, z, and $(1 \circ x) \circ w = x$ for all x whenever w is a power of two, so Definition 2.31 is satisfied. $\square$

An even shorter law (order 5) was obtained by the same author in a followup paper [**?**]:

Theorem 2.33 (Kisielewicz theoremII)**.** $(((1 \circ 1) \circ 1) \circ 0) \circ (1 \circ 2) \simeq 0$ *is an Austin law.*

Proof. Using the y^2 and y^3 notation as before, the law reads

$$(y^3 \circ x) \circ (y \circ z) = x. \tag{2.1}$$

In particular, for any y, the map $T_y \colon x \mapsto y^3 \circ x$ is injective, hence bijective in a finite model G. In particular we can find a function $f : G \to G$ such that $T_y f(y) = y^3$ for all y Applying (2.1) with $x = f(y)$, we conclude

$$T_y(y \circ z) = y^3 \circ (y \circ z) = f(y)$$

and thus $y \circ z$ is independent of z by injectivity of T_y. Thus, the left-hand side of (2.1) does not depend on x, and so the model is trivial. This shows there are no non-trivial finite models.

To establish an infinite model, use $\mathbb{N}$ with $x \circ y$ defined by requiring

$$y \circ y = 2^y; \quad 2^y \circ y = 3^y$$

and

$$3^y \circ x = 3^y 5^x$$

for $x \neq 3^y$, and

$$(3^y 5^x) \circ z = x$$

for $z \neq 3^y 5^x$. Finally set

$$2^{3^y} \circ z = 3^y$$

for $z \neq 3^y, 2^{3^y}$. All other assignments of $\circ$ may be made arbitrarily. It is then a routine matter to establish (2.1). $\square$

In that paper a computer search was also used to show that no law of order four or less is an Austin law.

Chapter 3

General implications

Definition 3.1 (Implication)**.** A law $w \simeq w'$ is said to *imply* another law $w'' \simeq w'''$, if every magma G that satisfies the former, satisfies the latter:

$$G \vDash w \simeq w' \implies G \vDash w'' \simeq w'''.$$

Two laws are said to be *equivalent* if they imply each other.

Lemma 3.2 (Pre-order)**.** *Implication is a pre-order on the set of laws, and equivalence is an equivalence relation.*

Proof. Trivial. □

Implications between these laws are depicted in Figure 3.1.

Lemma 3.3 (Minimal element)**.** *The law* $0 \simeq 0$ *is the minimal element in this pre-order.*

Proof. Trivial. □

Lemma 3.4 (Maximal element)**.** *The law* $0 \simeq 0$ *is the minimal element in this pre-order.*

Proof. Trivial. □

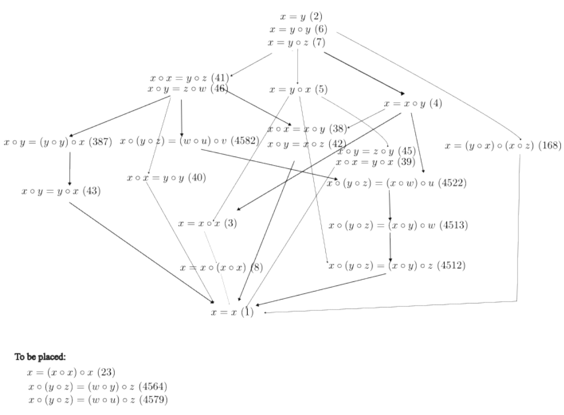

Figure 3.1: Implications between the above equations, displayed as a Hasse diagram.

Every magma G has a *reversal* G^{op}, formed by by replacing the magma operation $\circ$ with its opposite $\circ^{\mathrm{op}} : (x, y) \mapsto y \circ x$. There is a natural isomorphism between these magmas, which induces an involution $w \mapsto w^{\mathrm{op}}$ on words $w \in M_X$. Every law $w \simeq w'$ then has a *dual* $w^{\mathrm{op}} \simeq (w')^{\mathrm{op}}$.

For instance, the dual of the law $0 \circ 1 = 0 \circ 2$ is $1 \circ 0 = 2 \circ 0$, which after relabeling is $0 \circ 1 = 2 \circ 1$. A list of equations and their duals can be found here. Of the 4694 equations under consideration, 84 are self-dual, leaving 2305 pairs of dual equations.

The pre-ordering on laws has a duality symmetry:

Lemma 3.5 (Duality of laws)**.** *If $w \simeq w'$ implies $w'' \simeq w'''$, then $w^{\mathrm{op}} \simeq (w')^{\mathrm{op}}$ implies $w''^{\mathrm{op}} \simeq (w''')^{\mathrm{op}}$.*

Proof. This follows from the fact that a magma G satisfies a law $w \simeq w'$ if and only if G^{op} satisfies $w^{\mathrm{op}} \simeq (w')^{\mathrm{op}}$. □

Some equational laws can be "diagonalized":

Theorem 3.6 (Diagonalization)**.** *An equational law of the form*

$$F(x_1, \dots, x_n) = G(y_1, \dots, y_m), \tag{3.1}$$

where $x_1, \dots, x_n$ and $y_1, \dots, y_m$ are distinct elements of the alphabet, implies the diagonalized law

$$F(x_1, \dots, x_n) = F(x'_1, \dots, x'_n).$$

where $x'_1, \dots, x'_n$ are distinct from $x_1, \dots, x_n$ In particular, if $G(y_1, \dots, y_m)$ can be viewed as a specialization of $F(x'_1, \dots, x'_n)$, then these two laws are equivalent.

Proof. From two applications of (3.1) one has

$$F(x_1, \dots, x_n) = G(y_1, \dots, y_m)$$

and

$$F(x'_1, \dots, x'_n) = G(y_1, \dots, y_m)$$

whence the claim. □

Thus for instance, Definition 2.7 is equivalent to Definition 2.2.

Theorem 3.7 (Laws implied by the constant law)**.** *If w, w' each have order at least one, then the law $w \simeq w'$ is implied by the constant law (Definition 2.19). If exactly one of w, w' has order zero, and the law $w \simeq w'$ is not implied by the constant law.*

Proof. Routine. □

Theorem 3.8 (Criterion for implication)**.** *If $w \simeq w'$ is such that every variable appears the same number of times in both w and w', and $w \simeq w'$ implies another law $w'' \simeq w'''$, then every variable appears the same number of times in both w'' and w'''.*

Proof. Consider the magma $\mathbb{R}$ with the addition law $+$. By hypothesis, this magma obeys $w \simeq w'$, and hence $w'' \simeq w'''$, giving the claim by comparing coefficients of the linear forms associated to w'' and w''' in this magma. □

Chapter 4

Metatheorems from Invariants

For the purposes of this chapter, a *theorem* is a (true) statement about particular equations, for example '(387 implies 43)' is a theorem. A *metatheorem* is a general statement about implications; one can usually get many theorems from a single metatheorem. This chapter is all about generating many interesting metatheorems using a *meta-metatheorem*, called the fundamental property of invariants. If all this is making your head spin, don't worry. Look at [REFERENCES] for examples of metatheorems you can probably agree are both concrete and interesting. Once you have done that, come back here and we will show you how to prove these and other metatheorems using *invariants*.

4.1 Invariants

Let E, E_1, and E_2 be equations. If $E \Rightarrow E_1$ and $E_1 \Rightarrow E_2$, then $E \Rightarrow E_2$. Very trivial. Rephrasing this, we see that if $E \Rightarrow E_1$ and $E \not\Rightarrow E_2$, then $E_1 \not\Rightarrow E_2$.

Extending this idea, suppose we compute the set of all equations which are implied by E; we will call this set $\mathcal{Y}(E)$ (we use $\mathcal{Y}$ because this is an example of a Yoneda embedding). Then $\mathcal{Y}(E)$ is upwards closed, or closed under forward implication: no equation in $\mathcal{Y}(E)$ can imply an equation not in $\mathcal{Y}(E)$. If we know $\mathcal{Y}(E)$ well, this already settles a potentially large number of implications in the negative.

While computing $\mathcal{Y}(E)$ for an arbitrary equation E may seem daunting, for some nice equations we can find *invariants*, which makes the task manageable. An *invariant* for E is some sort of data associated with expressions w so that

$$\mathcal{Y}(E) = \{w = w' \mid \mathrm{Invariant}(w) = \mathrm{Invariant}(w')\}$$

If we can find an invariant which is computable for each term w, then we can easily describe $\mathcal{Y}(E)$. The fact that $\mathcal{Y}(E)$ is upwards closed is rephrased as follows; this is called **the fundamental property of invariants**. Remember that an invariant is a function taking expressions and outputting some data.

Meta-metatheorem 4.1 (Fundamental property of invariants)**.** *Let I be an invariant of E. If $w = w'$ implies $w'' = w'''$ and $I(w) = I(w')$ (that is, E implies $w = w'$), then $I(w'') = I(w''')$.*

More succinctly, for an invariant I of E we must have

$$(w = w' \Rightarrow w'' = w''') \implies (I(w) = I(w') \Rightarrow I(w'') = I(w''')).$$

When using this result, we commonly take the contrapositive: if $I(w) = I(w')$ and $I(w'') \neq I(w''')$, then $w = w'$ cannot imply $w'' = w'''$. Note that the conclusion is independent of the equation E; all we need to know is that I is an invariant.

Note for category theorists. Let Π denote the preorder of magma equations ordered by implication. If I is an invariant then define

$$I(w = w') := \begin{cases} \texttt{true} & \text{if } I(w) = I(w') \\ \texttt{false} & \text{otherwise} \end{cases}.$$

(In programming languages we would say $I(w = w') := I(w) == I(w')$). Let $\mathbf{Bool} = \{\texttt{true}, \texttt{false}\}$ be the poset where $\texttt{false} \leq \texttt{true}$. Then I becomes a function $\Pi \to \mathbf{Bool}$, and the fundamental property of invariants just says that this function is monotone, i.e. functorial. Thus for every invariant I we obtain a functor $\Pi \to \mathbf{Bool}$.

Question 1: Does every functor $\Pi \to \mathbf{Bool}$ come from an invariant?

Question 2: What can we say about the category of functors $\Pi \to \mathbf{Bool}$? Give a nice interpretation of natural transformations between invariants. $\square$

The fundamental property of invariants is not a theorem, nor a metatheorem: it is a meta-metatheorem, in the sense that it will allow us to get a metatheorem for every invariant we find.

Example: absorption law

Let E be the equation $x \circ y = x$. Intuitively, we must have

$$\mathcal{Y}(E) = \{w = w' \mid \text{the leftmost variable is the same for } w \text{ and } w'\}.$$

We will talk about proving statements like this one (say in Lean) later on; take it as given for now. The invariant is clear: we define $I(w)$ to be the leftmost variable of w. Instantiating this invariant in the fundamental property of invariants, we get the following metatheorem.

Metatheorem 4.2. *Let $w = w'$ be an equation such that the leftmost variable of w is the same as the leftmost variable of w'. Then $w = w'$ cannot imply an equation that does not have the property from the last sentence.*

Example: associativity

For a more complicated example, let E be the associativity equation $x \circ (y \circ z) = (x \circ y) \circ z$. Intuitively, we must have

$$\mathcal{Y}(E) = \{\text{equations that, when we remove all parentheses, are of the form } w = w\}.$$

There is an invariant lurking behind: it is the (ordered) list of variables appearing in an expression, counting repetitions. More formally, we define $I(w)$ to be the tuple of variables appearing in w, listed from left to right, say. Again, from the fundamental property of invariants we get the following.

Metatheorem 4.3. *Let $w = w'$ be an equation such that the variables appearing in w, taking into account order and repetitions, are the same ones that appear in w'. Then $w = w'$ cannot imply an equation that does not have the property from the last sentence.*

If we were coding a computer program that computes $I(w)$ given w, one could take the string of symbols that is w, ignore all parentheses, replace all symbols $\circ$ by commas, and surround with an appropriate delimiter. (I imagine one could easily do this using regular expressions.

We can compute other examples, but the invariant can get complicated even for simple equations. Exercise: what is the invariant for commutativity? Answer: To compute $I(w)$ from w replace all parentheses with curly braces and all symbols $\circ$ with commas, and interpret the result as nested sets.

4.2 Expanding the language

The method of invariants really shines when we expand our formal language. Right now our language consists of variables, parentheses, the equal sign, and $\circ$ (there is also an implicit use of $\forall$ but let's ignore that for now). Let Π denote the preorder of equations (built from the language described) ordered by implication.

We will add the symbol $\wedge$ ('and') to our language. Then we consider a bigger preorder $\Pi' \supseteq \Pi$ which includes equations and also conjunctions of equations. Even if we only care about Π it will be apparent that studying invariants in Π' gives us useful metatheorems about Π. Equations and conjunctions of equations are examples of *formulas* (or formulae, according to taste).

If φ is a formula, we can define $\mathcal{Y}(\varphi)$ to be the set of all formulae implied by φ; this agrees with our previous definition. Now define an invariant of φ to be a function I on terms such that

$$\mathcal{Y}(\varphi) \cap \Pi = \{w = w' \mid I(w) = I(w')\}.$$

Again, this clearly agrees with our previous definition. Although $\mathcal{Y}(\varphi) \cap \Pi$ might not be upwards closed in Π', it is upwards closed in Π, which is enough to get the fundamental property of invariants *verbatim*. This leads to more metatheorems we didn't have access to before.

Example: associativity and idempotency

Let φ be the conjunction of the associative law and the idempotency law ($x \circ x = x$). Again, we will rely on our intuition, which says that an invariant I defined by taking $I(w)$ to be the set of all variables appearing in w, works. The corresponding metatheorem is the following

Metatheorem 4.4. *Let $w = w'$ be an equation such that the set of variables appearing in w is equal to the set of variables appearing on w'. Then $w = w'$ cannot imply an equation that does not have the property from the last sentence.*

Example: associativity and commutativity

For a similar example, we can let φ be the conjunction of the associative and the commutative laws. Here we can define $I(w)$ to be the multiset of variables appearing in w. We obtain the following metatheorem.

Metatheorem 4.5. *Let $w = w'$ be an equation such that the variables appearing in w, taking into account multiplicity, are the same ones that appear in w'. Then $w = w'$ cannot imply an equation that does not have the property from the last sentence.*

Trivia: this was the first example of a metatheorem obtained by use of an invariant.

Example: associativity and commutativity with a twist

We can keep expanding our language if it helps us express more intricate invariants. For instance, we can add the symbol '1' to our language. Let φ be the conjunction of associativity, commutativity, the equations $1 \circ x = x$, and

$$\underbrace{x \circ x \circ \cdots \circ x}_{m \text{ times}} = 1,$$

for some fixed positive integer m. Pause to guess the invariant before we move on.

The invariant $I(w)$ is the multiset of variables appearing in w but multiplicities are computed modulo m. Thus we have the pretty metatheorem:

Metatheorem 4.6. *Fix some positive integer m. Let $w = w'$ be an equation such that every variable appearing in w appears the same number of times in w' modulo m. Then $w = w'$ cannot imply an equation that does not have the property from the last sentence.*

4.3 Proving metatheorems from invariants in Lean

For the rest of this chapter we readopt the convention of calling 'theorem' an important result, not necessarily pertaining to specific equations.

An invariant is generally a *syntactic* property of an expression. However, invariants can also be described and calculated *semantically* through the notion of a *lifting magma family*, described below. The general idea is that the value of an invariant for an expression can be computed by substituting specific values for the variables in the expression and evaluating the result in a certain magma in the lifting magma family; additional requirements ensure that the fundamental property of invariants is satisfied.

Definition 4.7 (Lifting Magma Family)**.** A *lifting magma family* is a family of magmas $\{G_\alpha\}$, one for each type α, satisfying the following properties:

- For each type α, there is a function $\iota_\alpha : \alpha \to G_\alpha$.
- Given a function $f : \alpha \to G_\alpha$, there is a magma homomorphism $\text{lift}\, f : G_\alpha \to G_\alpha$ such that $\text{lift}\, f(\iota_\alpha(x)) = f(x)$ for all x in α.

Example 4.8. The free Abelian groups form a lifting magma family. When the underlying set is finite, the groups are isomorphic to $\mathbb{Z}^n$.

Example 4.9. Lists form a lifting magma family.

The key consequence of the definition 4.7 is that it is significantly easier to check whether an equation is satisfied in a lifting magma family.

Theorem 4.10 (Evaluation theorem for lifting magma families)**.** *Suppose E is an equation involving a set of variables X, and let G be a lifting magma family.*

Determining whether E is satisfied by G_X is equivalent to checking that E is true with the specific substitution ι_X.

Proof. For the forward direction, suppose E is satisfied by G_X. Then, by definition, any substitution of the variables in E with elements of G_X will yield a true equation. In particular, substituting according to ι_X will yield a true equation.

For the reverse direction, suppose that E is true when evaluated with the substitution ι_X. Now, consider an arbitrary substitution of variables $f : X \to G_X$. By the lifting magma family property, there is a magma homomorphism $\operatorname{lift} f : G_X \to G_X$ such that $\operatorname{lift} f(\iota_X(x)) = f(x)$ for all x in X. In other words, applying the substitution f is equivalent to first applying to substitution ι_X and then applying the homomorphism $\operatorname{lift} f$. Since E is true when evaluated with the substitution ι_X, it is also true after applying the homomorphism $\operatorname{lift} f$. Thus, E is satisfied by G_X. □

Theorem 4.11 (The fundamental property of invariants)**.** *Let E and E' be equations involving a set of variables X, and let G be a lifting magma family.*

If E is true with the substitution ι_X, and E implies E', then so is E'.

Proof. Applying the evaluation theorem 4.10, we see that E is satisfied by G_X. Since E implies E', E' is also satisfied by G_X, and in particular, E' is true with the substitution ι_X. □

Remark 4.12. The result of evaluating an expression along the function $\iota_X : X \to G_X$ *is* the invariant.

In the case of Abelian groups, the result of evaluation is the variables in the expression with multiplicity. In the case of lists, the result of evaluation is the variables in the expression in the order they appear.

When the lifting magma family has good computational properties, calculating the invariant becomes easy.

Remark 4.13. Given an equation ϕ in the language of magmas (possibly involving logical operations other than equality and universal quantification), the initial (i.e., most general) magmas satisfying ϕ (provided they exist) form a lifting magma family.

However, for the purpose of generating invariants, we are interested in lifting magma families with convenient descriptions that are computationally tractable.

Remark 4.14. Suppose S is a finite set of equations in the language of magmas that is a confluent term rewriting system under a certain ordering of the terms (in the sense of the Knuth-Bendix algorithm). Then the initial magmas satisfying S form a lifting magma family where equality of elements in the magma is decidable.

This offers a way of generating examples of lifting magma families with good compuational properties for computing invariants of expressions.

4.4 Conclusion: Beyond Invariants

We are still lacking:

- A large collection of invariants.
- An estimate for how many implications the resulting metatheorems will settle.
- Algorithms (in Lean, Python, or otherwise) to compute known invariants.
- General results about lifting magmas.
- Formalization of the method of invariants and resulting metatheorems.
- Knowledge about the category-theoretic interpretation of invariants (see the questions in the note for category theorists).

Related to the last bullet point, we note the following. If all that matters about invariants is the fundamental property, we can apply the old French trick of turning a (meta-meta)theorem into a definition.

Q: If we were to define invariants as any functions satisfying the fundamental property, would anything change? (For those who read the note for category theorists: an equivalent redefinition is to consider invariants as functors $\Pi \to \mathbf{Bool}$).

Chapter 5

Subgraph implications

Interesting implications between the subgraph equations in Chapter 2. To reduce clutter, trivial or very easy implications will not be displayed here.

Theorem 5.1 (387 implies 43)**.** *Definition 2.22 implies Definition 2.17.*

Proof. (From MathOverflow). By Definition 2.22, one has the law

$$(x \circ x) \circ y = y \circ x. \tag{5.1}$$

Specializing to $y = x \circ x$, we conclude

$$(x \circ x) \circ (x \circ x) = (x \circ x) \circ x$$

and hence by another application of (2.22) we see that $x \circ x$ is idempotent:

$$(x \circ x) \circ (x \circ x) = x \circ x. \tag{5.2}$$

Now, replacing x by $x \circ x$ in (5.1) and then using (5.2) we see that

$$(x \circ x) \circ y = y \circ (x \circ x)$$

so in particular $x \circ x$ commutes with $y \circ y$:

$$(x \circ x) \circ (y \circ y) = (y \circ y) \circ (x \circ x). \tag{5.3}$$

Also, from two applications of (5.1) one has

$$(x \circ x) \circ (y \circ y) = (y \circ y) \circ x = x \circ y.$$

Thus (5.3) simplifies to $x \circ y = y \circ x$, which is Definition 2.17. ☐

Theorem 5.2 (29 equivalent to 14)**.** *Definition 2.11 is equivalent to Definition 2.9.*

This result was posed as Problem A1 from Putnam 2001.

Proof. By Lemma 3.5 it suffices to show that Definition 2.11 implies Definition 2.9. From Definition 2.11 one has

$$x = ((x \circ y) \circ x) \circ (x \circ y)$$

and also

$$y = (x \circ y) \circ x$$

giving $x = y \circ (x \circ y)$, which is Definition 2.9. ☐

Theorem 5.3 (29 equivalent to 14)**.** *Definition 2.11 is equivalent to Definition 2.9.*

This result was posed as Problem A1 from Putnam 2001.

Proof. By Lemma 3.5 it suffices to show that Definition 2.11 implies Definition 2.9. From Definition 2.11 one has

$$x = ((x \circ y) \circ x) \circ (x \circ y)$$

and also

$$y = (x \circ y) \circ x$$

giving $x = y \circ (x \circ y)$, which is Definition 2.9. □

The following result was problem A4 on Putnam 1978.

Theorem 5.4 (3744 implies 3722, 381)**.** *Definition 2.24 implies Definition 2.23 and Definition 2.21.*

Proof. By hypothesis, one has

$$x \circ y = (x \circ z) \circ (w \circ y)$$

for all x, y, z, w. Various specializations of this give

$$x \circ y = (x \circ z) \circ (y \circ y) \tag{5.4}$$

$$x \circ z = (x \circ z) \circ (x \circ z) \tag{5.5}$$

$$(x \circ z) \circ y = ((x \circ z) \circ (x \circ z)) \circ (y \circ y). \tag{5.6}$$

The equation (5.5) gives Definition 2.23, while (5.4), (5.5), (5.6) gives

$$x \circ y = (x \circ z) \circ y$$

which is Definition 2.21. □

Chapter 6

Subgraph counterexamples

Some counterexamples for the anti-implications between the subgraph equations in Chapter 2.

Theorem 6.1 (46 does not imply 4)**.** *Definition 2.19 does not imply Definition 2.4.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := 0$. ☐

Theorem 6.2 (4 does not imply 4582)**.** *Definition 2.4 does not imply Definition 2.30.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x$. ☐

Theorem 6.3 (4 does not imply 43)**.** *Definition 2.4 does not imply Definition 2.17.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x$. ☐

Theorem 6.4 (4582 does not imply 42)**.** *Definition 2.30 does not imply Definition 2.16.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y$ equal to 1 if $x = y = 0$ and 2 otherwise. ☐

Theorem 6.5 (4582 does not imply 43)**.** *Definition 2.30 does not imply Definition 2.17.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y$ equal to 3 if $x = 1$ and $y = 2$ and 4 otherwise. ☐

Theorem 6.6 (42 does not imply 43)**.** *Definition 2.16 does not imply Definition 2.17.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x$. ☐

Theorem 6.7 (42 does not imply 4512)**.** *Definition 2.16 does not imply Definition 2.25.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x + 1$. ☐

Theorem 6.8 (43 does not imply 42)**.** *Definition 2.17 does not imply Definition 2.16.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x + y$. ☐

Theorem 6.9 (43 does not imply 4512)**.** *Definition 2.17 does not imply Definition 2.25.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x \cdot y + 1$. ☐

Theorem 6.10 (4513 does not imply 4522)**.** *Definition 2.26 does not imply Definition 2.27.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y$ equal to 1 if $x = 0$ and $y \leq 2$, 2 if $x = 0$ and $y > 2$, and x otherwise. $\square$

Theorem 6.11 (4512 does not imply 4513)**.** *Definition 2.25 does not imply Definition 2.26.*

Proof. Use the natural numbers $\mathbb{N}$ with operation $x \circ y := x + y$. $\square$

Theorem 6.12 (387 does not imply 42)**.** *Definition 2.22 does not imply Definition 2.16.*

Proof. Use the boolean type Bool with $x \circ y := x || y$. $\square$

Theorem 6.13 (43 does not imply 387)**.** *Definition 2.17 does not imply Definition 2.22.*

Proof. Use the natural numbers $\mathbb{N}$ with $x \circ y := x + y$. $\square$

Theorem 6.14 (387 does not imply 4512)**.** *Definition 2.22 does not imply Definition 2.25.*

Proof. Use the reals $\mathbb{R}$ with $x \circ y := (x + y)/2$. $\square$

Theorem 6.15 (3 does not imply 42)**.** *Definition 2.3 does not imply Definition 2.16.*

Proof. Use the natural numbers $\mathbb{N}$ with $x \circ y := y$. $\square$

Theorem 6.16 (3 does not imply 4512)**.** *Definition 2.3 does not imply Definition 2.25.*

Proof. Use the natural numbers $\mathbb{N}$ with $x \circ y$ equal to x when $x = y$ and $x + 1$ otherwise. $\square$

Theorem 6.17 (46 does not imply 3)**.** *Definition 2.19 does not imply Definition 2.3.*

Proof. Use the natural numbers $\mathbb{N}$ with $x \circ y := 0$. $\square$

Theorem 6.18 (43 does not imply 3)**.** *Definition 2.17 does not imply Definition 2.3.*

Proof. Use the natural numbers $\mathbb{N}$ with $x \circ y := x + y$. $\square$

Chapter 7

Equivalence with the constant and singleton laws

85 laws have been shown to be equivalent to the constant law (Definition 2.19), and 815 laws have been shown to be equivalent to the singleton law (Definition 2.2).

These are the laws up to 4 operations that follow from diagonalization of 2.2 and of 2.19.

In order to formalize these in Lean, a search was run on the list of equations to discover diagonalizations of these two specific laws: equations of the form $x = R$ where R doesn't include x, and equations of the form $x \circ y = R$ where R doesn't include x or y.

The proofs themselves all look alike, and correspond exactly to the two steps described in the proof of 3.6. The Lean proofs were generated semi-manually, using search-and-replace starting from the output of `grep` that found the diagonalized laws.

In the case of the constant law, equation 2.15 ($x \circ x = y \circ z$) wasn't detected using this method. It was added manually to the file with the existing proof from the sub-graph project.

Chapter 8

Simple rewrites

53,905 implications were automatically generated by simple rewrites.

describe the process of automatically generating these implications here.

Chapter 9

Trivial auto-generated theorems

4.2m implications proven by a transitive reduction of 15k theorems were proven using simple rewrite proof scripts.

include more details of the methodology, and any comparisons with other generated implication data sets.