Associativity of Binary Operations on the Real Numbers

Samuel Joseph Audino
Bard College

Associativity of Binary Operations on the Real Numbers

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
of
Bard College

by
Samuel J. Audino

Annandale-on-Hudson, New York
May, 2016

Abstract

It is known that there is an agreed upon convention of how to go about evaluating expressions in the real numbers, denoted $\mathbb{R}$. We colloquially call this PEMDAS, which is short for Parentheses, Exponents, Multiplication, Division, Addition, Subtraction. It is also called the Order of Operations, since it is the order in which we execute the operators of a given expression. When we remove this convention and begin to execute the operators in every possible order, we begin to see that this allows for many different values based on the order in which the operations are executed. We will investigate this question by looking at how this affects the operations on $\mathbb{R}$ through using parentheses to force operators to be executed in a specific order. We compute the asymptotic bound for the number of outcomes, defined as associativity, for each of the operations on $\mathbb{R}$.

Contents

Dedication

I would like to dedicate this project to my mother, Kim, who has helped me find solace with who I am and who I have become, and to my father, Tom, who has been a great support to me in very trying times.

There is a select group of people who have meant very much to me for a very long time. While none of you helped me with my project, you've all helped me grow into a person capable of doing such things. Together we have all come into our own, and we will keep developing and continuing to do amazing things. You are my closest friends, and will continue to be. The Confed continues.

I would also like to dedicate this project to my close friends who I have made here, who have listened to me complain, and talk generally way too much. You are all wonderful people, whom I care so very much about.

Lastly, and most importantly, I would like to dedicate this to Lily and Garrett, for being there, anytime that I've needed them. You two are, and always will be, the best people I've known.

Acknowledgments

I would like to acknowledge Kline Commons for keeping me fed, for without the lovely dining hall I probably would have starved to death.

I would also like to acknowledge the Math department for accepting me into the program and making my three years here as enjoyable as possible.

1
An Introduction

1.1 PEMDAS and You

PEMDAS is a clever mnemonic used in schools to teach the order of operations on the real numbers, denoted $\mathbb{R}$, and its subsets (i.e., the rationals or the counting numbers). It is the arithmetic structure on the real numbers that describes the order in which operations are executed. Spelled completely out it is Parentheses, Exponentiation, Multiplication, Division, Addition, Subtraction. Even though multiplication is before division, and addition is before subtraction, they actually take the same precedent, and thus are just evaluated left to right. We also have the convention of evaluating exponentiation in a right to left manner. Using this convention we ensure that there is an unambiguous way to evaluate a given expression, which will then give us only one answer. This is fantastic because it removes the ambiguity from any expression you can write in $\mathbb{R}$. But a question arises from this: what would happen if we remove this structure? What happens if we execute these operations in other possible orders? Would we get the same answer, or would they differ wildly? Before we can tackle this question, we have to introduce some background information. We introduce the Catalan Numbers to gain access to a more insightful way of observing the results.

The Catalan Numbers were discovered in the early 18th century by the Mongolian mathematician Minggatu[2], but named after the man who applied them numerous counting problems, Eugéne Catalan[1]. The Catalan numbers are a sequence of numbers that are the solution to many different counting problems. We will focus on one of these problems in particular. We start with a simple idea: given n sets of parenthesis, what are all the possible ways in which they can be grouped such that the parentheses are proper? A proper set of parentheses is one where each right parenthesis '(' has a matching left parenthesis ')' when read left to right. For instance, '()((())())' is a set of proper parentheses, but '())(()(()(' is not, because the third, fourth, seventh and eighth parentheses are all unmatched.

Each of these pairs of parentheses can be viewed as a possible grouping of two numbers and a binary operation. When viewed this way, we can see that the set of different proper parentheses of length n for some $n \in \mathbb{N}$ is actually every possible way that an expression can be grouped. For instance, we see that an expression such as $a + b - c * d$ has three operators, and thus all of the balanced forms for 3 sets of parentheses are

$$((())), (())(),\ (()()),\ ()(()),\ ()()()$$

and those can be brought into correspondence with all possible groupings of the expression, listed below,

$$(a + (b - (c * d))), (a + ((b - c) * d)),\ ((a + b) - (c * d)),$$
$$((a + (b - c)) * d),\ (((a + b) - c) * d).$$

We will not go into how exactly this correspondence works, since it is not important for the project. It turns out that the set of different proper parentheses of length n for some $n \in \mathbb{N}$ are actually the well known Dyck words[1]. The Dyck Words are known to have

C_n different valid words[3], where C_n is the n-th Catalan Number. But before we start getting into any of this, we must first ask: what *are* the Catalan numbers?

1.2 The Catalan Numbers

We introduce the Catalan numbers below with their definition.

Definition 1.2.1. [3] Let $\{C_n\}_{n=0}^{\infty}$ be the sequence in $\mathbb{R}$ such that $C_0 = 1$ and $C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$. We call this sequence the **Catalan Numbers**, and C_n the n-th Catalan Number. $\triangle$

Using the above definition to evaluate the first few terms of the sequence we see the progression of the first few values is $\{1, 1, 2, 5, 14, \ldots\}$. The Catalan numbers are well studied, but we will still walk through some of the derivations in order to get a little more acquainted with them.

Our first investigation will be to find the closed formula for the n-th Catalan number, and we will use a recursive function in order to weasel our way to a solution.

Theorem 1.2.2. [3] *The closed formula of the n-th Catalan Number is*

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. Let $p\colon \mathbb{R} \to \mathbb{R}$ be a formal power series defined by $p(x) = 1 + x + 2x^2 + 5x^3 + \cdots = \sum_{n=0}^{\infty} C_n x^n$ where C_n is the n-th Catalan number. Note that $\sum_{k=0}^{n} C_k C_{n-k-1} = C_n$. Therefore, we have

$$(p(x))^2 = \left(\sum_{n=0}^{\infty} C_n x^n\right)\left(\sum_{n=0}^{\infty} C_n x^n\right) \tag{1.2.1}$$

$$= \sum_{n=0}^{\infty}\left[\left(\sum_{k=0}^{n} C_k C_{n-k}\right) x^n\right]. \tag{1.2.2}$$

By our definition, $\sum_{k=0}^{n} C_k C_{n-k} = C_{n+1}$. Therefore, we have that the above expression reduces to $\sum_{k=0}^{n} C_{n+1} x^n$. Note that this is nearly our original function $p(x)$. We further multiply both sides of the equation by x, giving

$$x[p(x)]^2 = \sum_{k=0}^{n} C_{n+1}x^{n+1} = p(x) - 1 \tag{1.2.3}$$

$$p(x) = x[p(x)]^2 + 1 \tag{1.2.4}$$

$$p(x) = \frac{1 - \sqrt{1-4x}}{2x}. \tag{1.2.5}$$

We can use the extended Binomial Theorem we can expand the part under the radical,

$$\sqrt{1-4x} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n}(-4x)^n = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{4^n(2n-1)}\binom{2n}{n}(-1)^n 4^n x^n = \sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{2n-1}\binom{2n}{n}x^n. \tag{1.2.6}$$

Plugging the radical into $p(x)$, we get

$$p(x) = \frac{1-\sqrt{1-4x}}{2x} = \frac{1 - \sum_{n=0}^{\infty} \frac{(-1)^{2n+1}}{2n-1}\binom{2n}{n}x^n}{2x} \tag{1.2.7}$$

$$= \frac{1 - 1 - \sum_{n=1}^{\infty} \frac{-1}{2n-1}\binom{2n}{n}x^n}{2x} \tag{1.2.8}$$

$$= \sum_{n=0}^{\infty} \binom{2n}{n}\frac{1}{n+1}x^n. \tag{1.2.9}$$

Since the coefficient for each x^n is C_n, then $C_n = \binom{2n}{n}\frac{1}{n+1}$. □

We now have the closed form for the value of C_n is for any $n \in \mathbb{N}$, and we will continue to parse through some of the undefined terms from the first section by introducing the definition for Dyck words.

Definition 1.2.3. [1] A **Dyck Word** is a string of length $2n$ containing X's and Y's that satisfies the following criteria:

1. The string must contain n X's and n Y's.

2. No initial segment of the word can contain more Y's than X's.

△

Dyck words are extremely helpful for looking at certain strings of characters. For example, the set of Dyck words of length 6 would be $\{XXXYYY,\ XXYXYY,\ XYXXYY,\ XYXYXY,\ XXYYXY\}$. Note that Dyck words always start with an X and end with a Y, and that the set of Dyck words of length $2n$ is C_n. Since the X's and Y's are arbitrary, we reinterpret them in the following way: we'll define X and '(' and Y and ')'. It is easy to see that using those terms the Dyke words are in equivalence with the set of balanced parentheses. For example, $XXYXYY$ is equivalent to $(()())$. As previously mentioned, is directly related to the ways in which we can parenthesize an expression with n operators. We can see these words as pairing two objects with a binary operation. Returning to the example we introduced at the beginning of the paper, we see that the balanced parenthesized forms of an expression such as $a + b - c * d$ are

$$(a + (b - (c * d))),\ (a + ((b - c) * d)), ((a + b) - (c * d)),$$
$$((a + (b - c)) * d), (((a + b) - c) * d).$$

We can view each of these pairings as a distinct function. Also note that something that is parenthesized and contains binary operations is an object in and of itself. This means that if we have $(a+(b-(c*d)))$, then the object 'pairs' are $c*d$, $b-(c*d)$, and $a+(b-(c*d))$. We can write all the possible orderings of the expression as what we call a Tamari Lattice[4], pictured on the next page.

We will describe the Tamari Lattice in greater detail later. Since these are all of the possible different ways in which to execute the binary operations, we have the following result:

Theorem 1.2.4. *The upper bound for possible unique answers for an expression with n operators is C_n.*

We will take a moment here to talk about what associativity is. We will define this relative associativity by looking at how many different answers the expression gives. For

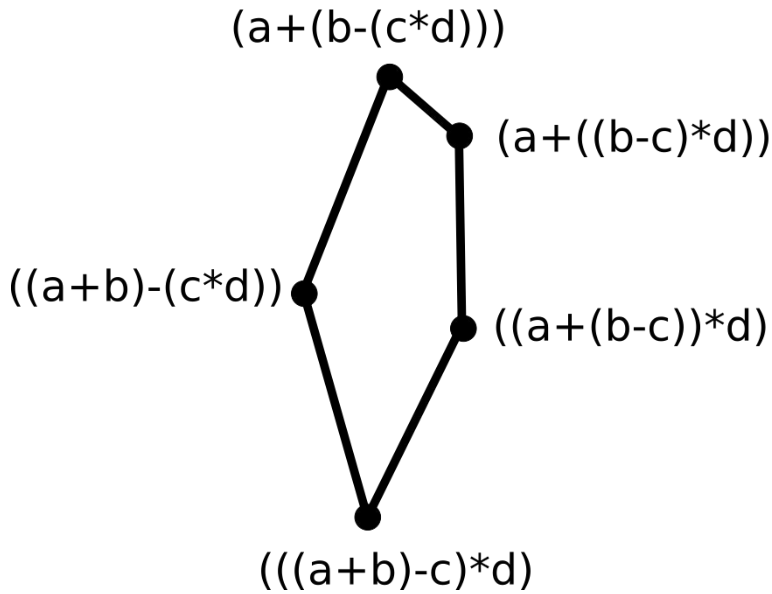

Figure 1.2.1. The Tamari Lattice for the expression $a + b - c * d$. See section 3.2 for a introduction to the lattice structure.

instance, we call an expression **completely associative** if there is only one distinct answer no matter the order in which the operations of the expression are evaluated. An expression is **completely unassociative** if there are C_n distinct answers depending on the order in which the operations are evaluated. One expression is more associative than another if there are less distinct answers.

2

Operations on $\mathbb{R}$

In this chapter we will investigate five operations on $\mathbb{R}$: addition, subtraction, multiplication, division, and exponentiation. We must begin by setting up some formal definitions first, and then we will continue with what it means to be associative. We will see that addition and multiplication are completely associative, while subtraction, division, and exponentiation are not as fortunate.

We introduce the general form for the expressions that we are going to talk about.

2.1 n-expressions

Before we begin talking about anything involving expressions, we must define what we've meant by the word *expressions*.

Definition 2.1.1. An ***n*-expression** in $\mathbb{R}$ is a string S of $n+1$ real numbers $\{a_1, a_2, \ldots, a_{n+1}\}$ and n operators $\{*_1, *_2, \ldots, *_n\}$ written in the following way,

$$a_1 *_1 a_2 *_2 \cdots *_n a_{n+1}. \tag{2.1.1}$$

If all the operators are of the same type, we call the expression a (name of the operator) n-expression. △

There is another type of expression we need to define is called a **normal form**.

Definition 2.1.2. We define a **normal form** of an n-expression in $\mathbb{R}$ to be an n-expression containing no parentheses, and following the typical order of operations evaluated on $\mathbb{R}$. △

For instance, $a_1 *_1 a_2 *_2 a_3 *_3 a_4$ is a "normal form." The difference between an n-expression and a normal form is that former can have parentheses pairing the numbers while the latter cannot. We introduce these because they are incredibly helpful for comparing different parenthesized expressions in $\mathbb{R}$. We must also introduce a helpful way of organizing all of the different parenthesized expressions, which is the Tamari Lattice structure.

2.2 Tamari Lattices

The Tamari Lattice structure, construced by Dov Tamari in 1968[4], is a poset wherein the elements of the lattice are groupings of objects into pairs via parentheses. As we saw earlier in Figure 1.2.1, we can take the set of parenthesized expressions to construct a Tamari Lattice. We define the lattice in the following way.

Definition 2.2.1. A **Tamari Lattice** is a partially ordered set wherein the elements of the lattice are groupings by pairs of objects with parentheses. The ordering on the lattice is such that one element is greater than the other if and only if the greater element can be obtained from the lesser by only rightward applications of the associative law. △

For instance (from our previously introduce lattice in Figure 1.2.1), $(a+((b-c)*d)) > ((a+(b-c))*d)$, since the right handed side is equal to the left handed side after one rightward application of the associativity law. This is an incredibly helpful structure since we can now have a visual for what these parenthesized expressions are.

We will also define the following two sets from the elements of the Tamari Lattice. Given a set of numbers $\{a_1, a_2, \ldots, a_{n+1}\}$ and operators $\{*_1, *_2, \ldots, *_n\}$ written as a normal expression, we can define the following sets.

Definition 2.2.2. Let $n \in \mathbb{N}$. Let $F^n = \{S_1, \ldots, S_{C_n}\}$ be a set of functions where each $S_i \colon \mathbb{R}^{n+1} \to \mathbb{R}$ for $1 \leq i \leq n$ corresponds to an element of the Tamari Lattice. We define F^n as the function space of all parenthesized versions of an n-expression. $\triangle$

Note that $|F^n| \leq C_n$ always, since there are C_n elements of the Tamari Lattice of an n-expression. Thus if all the S_i are distinct functions, then we have $|F^n| = C_n$. We also want to define a set of normal forms, in order to show that we can map this function space onto this second set of functions.

Definition 2.2.3. Let $n \in \mathbb{N}$. Let $N^n = \{N_1, \ldots, N_k\}$ for some $k \in \mathbb{N}$, be a set of functions where each $N_i \colon \mathbb{R}^{n+1} \to \mathbb{R}$ for $1 \leq i \leq k$ corresponds to some normal form of the n-expression. We call this the set of possible normal forms of an expression with n operators. $\triangle$

The magnitude of N^n depends on the expression, and will be established in each case. We will show why the associativity of the expression depends on the magnitude of N^n by showing that F^n maps onto N^n via the underlying arithmetic structure of whatever ring we are working in. We can thus define associativity in the following way.

Definition 2.2.4. The associativity of an expression is defined as the magnitude of N^n. If $|N^n| = 1$, then the expression is **completely associative**. If $|N^n| = C_n$, then the expression is **completely unassociative**. If there are two expressions, then one expression is more associative than the other if there are less normal forms for that expression. $\triangle$

2.3 Addition and Multiplication on $\mathbb{R}$

Addition and multiplication have similar properties in $\mathbb{R}$. They are both associative, which we will show means that they are also completely associative. We will still investigate addition, and make a brief remark about multiplication, to outline the methodology for determining the associativity of an operator. We start with the Tamari Lattices generated by general addition n-expressions. We saw that for any n-expression there existed a Tamari Lattice structure in which each node was a different way of executing an n-expression, and if we treat the nodes of the lattice as functions $S_1, \ldots, S_{C_n} : \mathbb{R}^{n+1} \to \mathbb{R}$. Let us take the following addition n-expression,

$$a_1 + a_2 + a_3 + a_4$$

We will set up the Tamari Lattice for this expression,

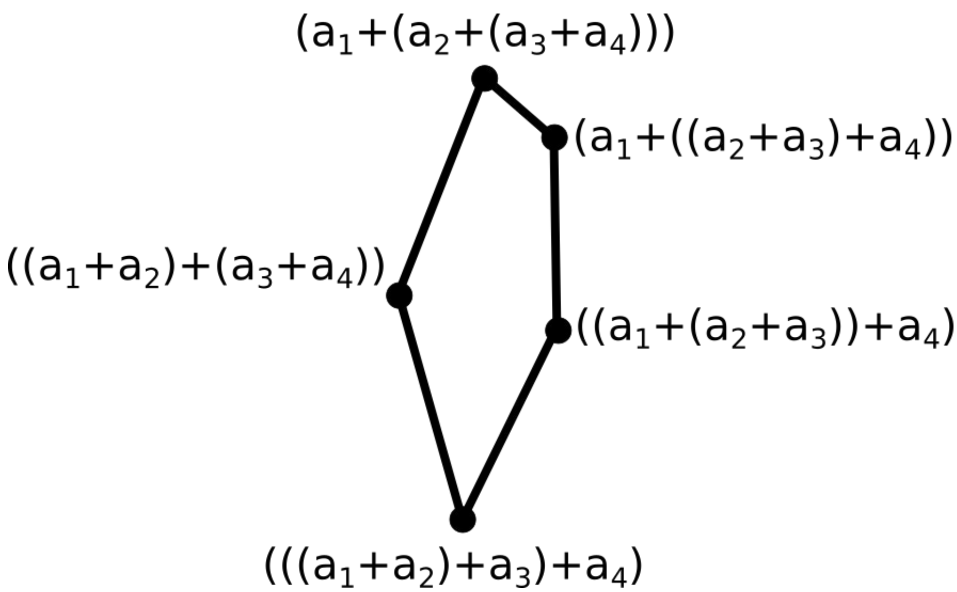

Figure 2.3.1. The Tamari Lattice for the expression $a_1 + a_2 + a_3 + a_4$.

By applying the structure of $\mathbb{R}$ to remove the parentheses from each of the functions in F^3, we can see that all the elements of F^3 map to the single element of N^3. The only normal form of that expression is just $a_1+a_2+a_3+a_4$, i.e., that $N^3 = \{a_1+a_2+a_3+a_4\}$.

This means that all the different functions defined by the elements of the Tamari Lattice are actually the same, and thus there is only one distinct function present. Therefore it is completely associative by our definition.

This leads us to our first result,

Theorem 2.3.1. *The addition n-expression in $\mathbb{R}$ is completely associative.*

Proof. Let $n \in \mathbb{N}$ and $a_1, \ldots, a_{n+1} \in \mathbb{R}$. The addition n-expression is written as $a_1 + \cdots + a_{n+1}$. Note that there is only one possible normal form of this expression, i.e., $N^n = \{a_1 + \cdots + a_{n+1}\}$. If we apply the transformation via the structure of $\mathbb{R}$, we can see that each parenthesized expression reduces to our one element of N^n. Thus F^n maps onto the one element of N^n, and therefore all the C_n elements of the Tamari Lattice reduce to the same function, namely, the one normal form for the addition n-expression. $\square$

This should be of no surprise, since we already know that addition is associative in $\mathbb{R}$, which means one can remove and add parentheses without worry of changing the expression. By a similar logic, it is obvious that any associative operator in $\mathbb{R}$ must also be completely associative in $\mathbb{R}$.

Theorem 2.3.2. *The multiplication n-expression in $\mathbb{R}$ is completely associative.*

2.4 Subtraction on $\mathbb{R}$

Subtraction requires more of an investigation of how these normal forms are derived. We can observe the Tamari Lattices generated by general subtraction n-expressions in order to see how exactly why we choose what the normal forms for subtraction are. We saw that for any n-expression there existed a Tamari Lattice structure in which each node was a different way of executing an n-expression. When we treat the nodes of the lattice as functions $S_1, \ldots, S_{C_n} : \mathbb{R}^{n+1} \to \mathbb{R}$ we can create an algorithm to check whether or not the functions are unique within the set of nodes of the Tamari Lattice.

Looking at any node we can notice two very important facts about each subtraction n-expression. For any subtraction n-expression, we know that a_1 is positive, and a_2 is always subtracted from a_1. This is due to the underlying structure on $\mathbb{R}$, that allows us to distribute negative signs in order to remove parentheses. We want to use this fact to convert these parenthetical expressions into their normal forms.

For example, the normal form of the subtraction 2-expression $(a_1 - (a_2 - a_3))$ is $a_1 - a_2 + a_3$, using the arithmatic structure on the ring $\mathbb{R}$, while the normal form of the other possible subtraction 2-expression $((a_1 - a_2) - a_3)$ is $a_1 - a_2 - a_3$. Before we continue we must stress again the important fact: for any subtraction n-expression, the normal form of that expression always begins as $a_1 - a_2 \cdots$. This will help us tremendously, as it will reduce the amount of possible normal forms of subtraction by an entire factor of two. We will introduce an algorithm that reverses this process, showing that each normal form corresponds with a certain parenthesized subtraction n-expression.

Lets take a minute to truly clarify the normal forms of subtraction. The set of normal forms N^n for subtraction consist of elements written in the following way

$$a_1 - a_2 \pm a_3 \pm a_4 \pm \cdots \pm a_{n+1}.$$

This is because the distribution of negative signs via removing parentheses can cause any operator after the first to be either an addition or subtraction operator. Since each other operator can either be an addition or subtraction, then there must be 2^{n-1} different normal forms. Therefore,

Lemma 2.4.1. *For an subtraction n-expression, $|N^n| = 2^{n-1}$.*

2.5 Two Algorithms for Subtraction

We will show via two algorithms that not only does each normal form correspond with a parenthesized n-expression, but that a single normal form can map to different parenthesized n-expressions based on which algorithm you use. The first algorithm will show that we can map these normal forms back to the parenthesized expressions, and then the second algorithm will show us that not every parenthesized expression has a unique normal form for subtraction. Let us introduce the first algorithm that we will be using in order to convert normal forms into parenthesized n-expressions.

Definition 2.5.1. Let $n \in \mathbb{N}$. Let $\mathcal{A}\colon N^n \to F^n$ be a function defined in the following way,

1. If we have a negative term followed by a positive one (ie, $\cdots - a_\alpha + a_{\alpha+1} \cdots$) we factor out a negative one, giving a parenthesized version of those two objects (ie, $\cdots - (a_\alpha - a_{\alpha+1}) \cdots$).
2. If there are no more addition signs, we left associate the terms beginning (ie, $a_1 - a_2 - a_3 - \cdots \to ((a_1 - a_2) - a_3) - \cdots$).

$\triangle$

Lets introduce the Tamari Lattice for a subtraction 3-expression for our example. It is pictured below,

To exemplify how this algorithm works, let us look at all the normal forms of an 3-expression using only addition and subtraction operators. We list out all of these expressions below

$$a_1 - a_2 - a_3 - a_4,\ a_1 - a_2 - a_3 + a_4,\ a_1 - a_2 + a_3 - a_4,\ a_1 - a_2 + a_3 + a_4$$

$$a_1 + a_2 - a_3 - a_4,\ a_1 + a_2 - a_3 + a_4,\ a_1 + a_2 + a_3 - a_4,\ a_1 + a_2 + a_3 + a_4.$$

Using our noted fact, we can throw away those normal forms that start $a_1 + a_2$, since they will never be the result of any of our parenthesized subtraction 3-expressions. This leaves us with 4 different expressions,

$$a_1 - a_2 - a_3 - a_4,\ a_1 - a_2 - a_3 + a_4,\ a_1 - a_2 + a_3 - a_4,\ a_1 - a_2 + a_3 + a_4.$$

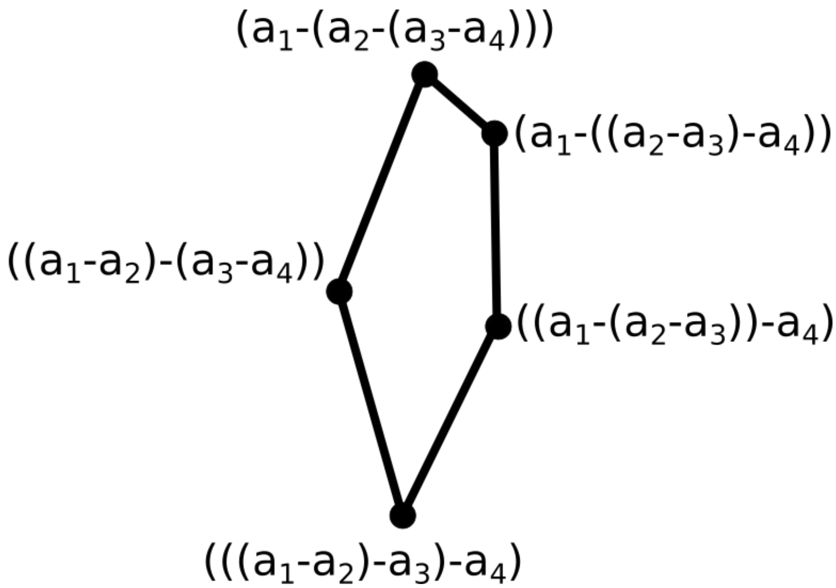

Figure 2.5.1. The Tamari Lattice for the expression $a_1 - a_2 - a_3 - a_4$.

Let us denote the set of such expressions as N^3. We will pick $a_1 - a_2 + a_3 - a_4$ as an exemplary member to show how the algorithm works. Applying the first rule,

$$a_1 - a_2 + a_3 - a_4 \quad \rightarrow \quad a_1 - (a_2 - a_3) - a_4$$

wherein we must apply rule number 2,

$$\rightarrow \quad (a_1 - (a_2 - a_3)) - a_4$$

and then rule number 2 again,

$$\rightarrow \quad ((a_1 - (a_2 - a_3)) - a_4)$$

which fully parenthesizes our expression.

Showing that the other three 3-expressions can be put through $\mathcal{A}$ in order to get different parenthesized subtraction 3-expressions will be left as an exercise for the reader in order to get them more acquainted with the algorithm.

We will also introduce another algorithm in order to show that some nodes of the Tamari Lattice are redundant.

Definition 2.5.2. Let $\mathcal{B}\colon N^n \to F^n$ be a function defined in the following way,

1. If an addition is followed by a subtraction (i.e., $\cdots + a_\beta - a_{\beta+1} \cdots$) we can parenthesize the two terms together, without factoring out (-1), (i.e., $\cdots + (a_\beta - a_{\beta+1}) \cdots$).
2. If we have a negative term followed by a positive one (i.e., $\cdots - a_\alpha + a_{\alpha+1} \cdots$) we factor out a (-1), giving a parenthesized version of those two objects (i.e., $\cdots - (a_\alpha - a_{\alpha+1}) \cdots$).
3. If there are no more addition signs, we left associate the terms beginning (i.e., $a_1 - a_2 - a_3 - \cdots \to ((a_1 - a_2) - a_3) - \cdots$).

$\triangle$

Applying $\mathcal{B}$ the same expression we applied $\mathcal{A}$ to, we can see that

$$\begin{aligned} a_1 - a_2 + a_3 - a_4 &\to a_1 - a_2 + (a_3 - a_4) \\ &\to a_1 - (a_2 - (a_3 - a_4)) \\ &\to (a_1 - (a_2 - (a_3 - a_4))). \end{aligned}$$

This shows us that some of the nodes on this Tamari Lattice actually define the same function, and thus are redundant. Therefore we conjecture there must be fewer than C_n possible results for any subtraction n-expression.

This points us towards our next result, that there actually are only 4 distinct functions of a Tamari Lattice of $a_1 - a_2 - a_3 - a_4$. Since there are only 2^{n-1} possible normal forms for any parenthesized subtraction n-expression, this directly implies that there at most 2^{n-1} unique functions from $\mathbb{R}^{n+1} \to \mathbb{R}$ for the normal forms of subtraction with n-operators. We generalize these results below.

Lemma 2.5.3. *If $a_1, \ldots, a_{n+1} \in \mathbb{R}$ for some $n \in \mathbb{N}$. Let $A = \{a_i \,|3 \le i \le n+1\}$ be a set. If A satisfies the following criteria,* *1. If C_1, C_2 are subsets of A and $C_1 \neq C_2$ then $\sum C_1 \neq \sum C_2$,*

then all the possible normal forms of the subtraction n-expression for this selection of numbers are unique functions.

Proof. Before we begin, by our hypothesis it follows that $a_3, \ldots, a_{n+1} \neq 0$. Let A be a set as defined in our hypothesis, with $a_1, \ldots, a_{n+1} \in \mathbb{R}$. Let $C_1, C_2 \subseteq A$, let $C_2 = \emptyset$, and $C_1 \neq \emptyset$. The sum of an empty subset is defined as $\sum \emptyset = 0$. Then by our criteria, we know that $\sum C_1 \neq \sum C_2 = 0$. Therefore, no subset of A sums to zero.

We will now show that under these conditions each normal form must be distinct. Let N_i, N_j be two different elements of N^n, where $i, j \in \mathbb{N}$ and $1 \leq i, j \leq n$. Assume that $N_i = N_j$. Since they are two different elements of N^n, their signs must differ at some amount of a_k for $3 \leq k \leq n+1$.

Let us define two sets, C_1, C_2, where C_1 is all the a_k that are positive in N_i and negative in N_j, and C_2 is all the a_k that are negative in N_i and positive in N_j. Notice that once the like terms are canceled, we can move all the elements of C_1 in N_j to the left hand side, and similarly can move all the elements of C_2 in N_i to the right hand side. This gives $2\sum C_1 = 2\sum C_2$. Notice that $C_1, C_2 \subseteq A$ and $C_1 \neq C_2$. Therefore, by our criteria, we have a contradiction.

$\square$

Since for our choice of numbers all the normal forms of an subtraction n-expression are unique, therefore by lemma 2.4.1 and 2.5.3 we have the following result.

Theorem 2.5.4. *The lower bound for distinct answers for a subtraction n-expression is 2^{n-1}.*

This comes about from the two parts of our analysis: that we can map F^n onto N^n using the arithmetic structure on $\mathbb{R}$, and that $|N^n| = 2^{n-1}$ under our specific conditions. Together this implies the above result.

2.6 Division on $\mathbb{R}$

The analysis of the associativity of division is very similar to subtraction. Let us take an arbitrary division n-expression, denoted $a_1 \div a_2 \div \cdots \div a_{n+1}$. Looking at the elements of the Tamari Lattice we can see that every expression begins with a_1 times $\frac{1}{a_2}$. This means the normal forms of a division n-expression can be written as fractions, with a factor of a_1 in the numerator and a_2 in the denominator, and the rest of the elements either being in the denominator or the numerator. Therefore we have two different states each number a_3 to a_{n+1} can be and therefore 2^{n-1} possible variations.

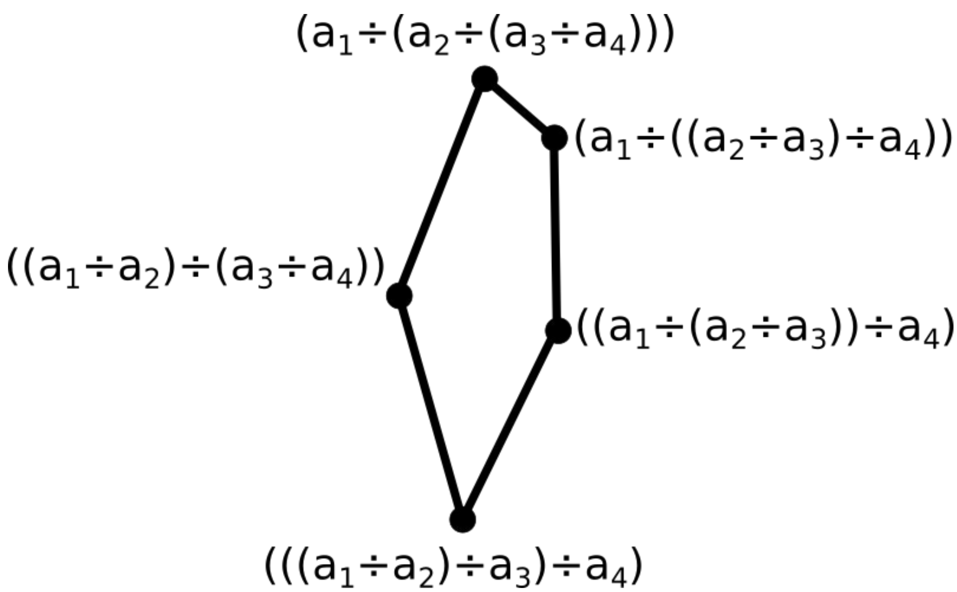

Figure 2.6.1. The Tamari Lattice for the expression $a_1 \div a_2 \div a_3 \div a_4$.

For example, take the normal division n-expression $a_1 \div a_2 \div a_3 \div a_4$. Looking at the Tamari Lattice for this structure we can see, using the arithematic structure of $\mathbb{R}$, that $a_1 \div a_2$ will always be the case. Otherwise we have that a_3 and a_4 can be placed either in the denominator (dividing $\frac{a_1}{a_2}$) or in the numerator (multiplying $\frac{a_1}{a_2}$). Thus the normal forms of a division 3-expression are,

$$\frac{a_1a_3a_4}{a_2}, \frac{a_1a_3}{a_2a_4}, \frac{a_1a_4}{a_2a_3}, \frac{a_1}{a_2a_3a_4}.$$

With this realization, we generalize to get the following result.

Lemma 2.6.1. *For a division n-expression, $|N^n| = 2^{n-1}$.*

Using the arithmetic structure on $\mathbb{R}$, obtain a map from the elements of the Tamari Lattice for a division n-expression F^n onto the normal forms of that expression N^n. We once again introduce constraints in order to show that it is possible to pick numbers $a_3, \cdots, a_{n+1}$ such that all of these normal forms give distinct values.

Lemma 2.6.2. *If $a_1, \ldots, a_{n+1} \in \mathbb{R} - \{0\}$ for some $n \in \mathbb{N}$. Let $A = \{a_i \mid 3 \leq i \leq n+1\}$ be a set. If A satisfies the following criteria,*

1. *If C_1, C_2 are subsets of A and $C_1 \neq C_2$, then $\Pi C_1 \neq \pm \Pi C_2$.*

then all the possible normal forms of division are unique functions.

Proof. Before we begin, by our hypothesis it follows that $a_1, \ldots, a_{n+1} \neq \pm 1$. Let A be a set as defined in our hypothesis, with $a_1, \ldots, a_{n+1} \in \mathbb{R} - \{0\}$. We disallow 0 in order to keep our calculations well defined. Let $C_1, C_2 \subseteq A$, let $C_2 = \emptyset$, and $C_1 \neq \emptyset$. The multiplicitive sum of an empty subset is defined as $\Pi\emptyset = 1$. Then by our criteria, we know that $\Pi C_1 \neq \Pi C_2 = 1$. By our criteria, we also have that $\Pi C_1 \neq -\Pi C_2 = -1$. Therefore, no non-empty subset of A mutliplicatively sums to ± 1.

We will now show it is impossible for two normal forms to be equal under these conditions. Let $n \in \mathbb{N}$ and $N_i, N_j \in N^n$ be two distinct normal forms for $i, j \in \mathbb{N}$ and $i, j \leq 2^{n-1}$. Assume $N_i = N_j$. If the two normal forms are distinct, then they must have a certain number of a_k for $k \in \mathbb{N}$ that are in the denominator in N_i but in the numerator in N_j, or vice versa. Let C_1 be the set of all a_k in the numerator of N_i and also in the denominator of N_j, and let C_2 be the set of all a_j in the denominator of N_i and also in the numerator of N_j. Note that $C_1, C_2 \subseteq A$, that $C_1 \neq C_2$ and if we divide both sides of our assumed equality by N_j, we get,

$$\frac{N_i}{N_j} = \frac{(\Pi C_1)^2}{(\Pi C_2)^2} = 1$$
$$\frac{\Pi C_1}{\Pi C_2} = \pm 1.$$

If the above expression were equal to ± 1, it would be a contradiction, since a non-empty subset of A would multiplicatively sum to ± 1.

$\square$

And therefore by lemma 2.6.1 and 2.6.2,

Theorem 2.6.3. *The lower bound for distinct answers for a division n-expression is 2^{n-1}.*

Interestingly enough, this means that both division and subtraction are equally associative in $\mathbb{R}$. This shouldn't be much of a surprise, since we have seen that the magnitude of the set of normal forms, N^n, is the indicator of how associative an expression can be.

2.7 Exponentiation on $\mathbb{R}$

Exponentiation is where further analysis is required. Unlike the other operations on $\mathbb{R}$, exponentiation is not as straight forward. There is less intuition as to which expressions with a repeated exponentiation operator are equivalent. We will show that the maximum amount of normal forms is actually C_n for $n = 2, 3, 4$. We wish to start with an example to denote the normal forms of exponentiation, and to highlight this difficulty.

Let $a_1, a_2, a_3 \in \mathbb{R} - \{0\}$. We disallow 0 because of problems involving the term 0^0, since we would like our expressions to always be well defined.

We want to apply the structure on $\mathbb{R}$ to remove the parentheses to get the normal forms of these expressions, which are

$$a_1^{a_2^{a_3}},\ a_1^{a_2 a_3}.$$

Note that a_1 is always the base. Therefore, we can make these normal forms a bit neater by applying log_{a_1} to both sides. This gives us the simpler normal forms for exponentiation

$$a_2^{a_3},\ a_2 a_3.$$

We want to look at the solution set of

$$a_2^{a_3} - a_2 a_3 = 0, \tag{2.7.1}$$

and we must ask, when are these two normal forms equal? There is the simple solution of $a_2 = a_3 = 1$, but are there other solutions? Choosing only $a_3 = 1$, we can see that a_2 can be any value, since the normal forms reduce from $a_2^{a_3} = a_2 a_3$ to $a_2 = a_2$. Another easily observed solution is $a_2 = a_3 = 2$. We can ask the honest question, does this also lie on a curve or is it an isolated solution? Since 2.7.1 is a differentiable function and that a_3 is defined as an implicit function of a_2, we know that $a_2 = a_3 = 2$ is not an isolated solution. So there exists two curves that allow those two to be equivalent.

Since we are looking for when they are not, this gets a bit easier for us. Let $a_2 = a_3 = \frac{1}{2}$. Then $a_2^{a_3} = \sqrt{\frac{1}{2}}$ and $a_2 a_3 = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$, which are obviously different.

Now let $a_1, a_2, a_3, a_4 \in \mathbb{R} - \{0\}$. The Tamari Lattice for this expression is pictured on the next page, and this gives us the normal forms,

$$a_2^{a_3^{a_4}},\ a_2^{a_3 a_4},\ a_4 a_2^{a_3},\ a_2 a_3^{a_4},\ a_2 a_3 a_4.$$

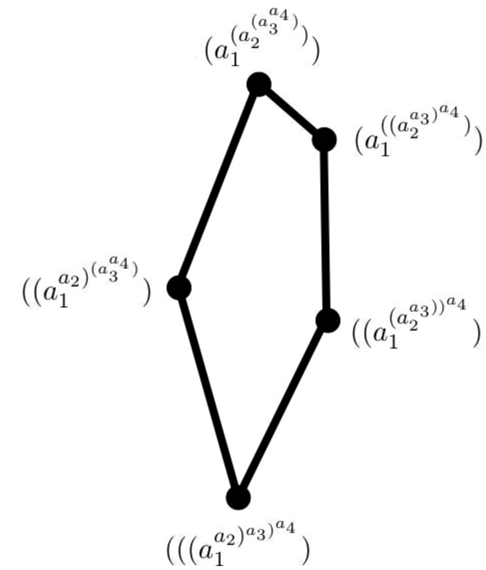

Figure 2.7.1. The Tamari Lattice for the exponentiation 3-expression $a_1^{a_2^{a_3^{a_4}}}$.

It is very easy to see how the number of functions to check quickly increases. However, this allows us to realize something important about these functions. If we avoid choosing numbers $a_1, \ldots, a_4$ such that $a_2 = a_3 = a_4$, these are all distinct functions. Looking at the third and fourth normal forms for this expression let us realize that we must avoid the situation wherein $a_2^{a_3} = a_3^{a_4}$ in order to produce the maximum amount of normal forms. So let us choose $a_1 = \frac{1}{2}, a_2 = \frac{1}{2}, a_3 = \frac{1}{4}, a_4 = \frac{1}{8}$. Then the normal forms evaluate to,

$$2^{-\frac{1}{\sqrt[4]{2}}},\ \frac{1}{\sqrt[32]{2}},\ \frac{1}{8\sqrt[4]{2}},\ \frac{1}{2\sqrt[8]{4}},\ \frac{1}{64}.$$

Therefore we have that each of these are distinctly different numbers, and thus all the normal forms are distinct.

The problem is in finding the correct numbers to choose in this situation. We will show for $n = 4$ that the selection of numbers $a_1 = \frac{1}{2}, a_2 = \frac{1}{4}, a_3 = \frac{1}{8}, a_4 = \frac{1}{16}, a_5 = \frac{1}{32}$ that the fourteen normal forms of the exponentiation 4-expression are unique.

The set of functions F^4 can be written,

$$F^4 = \{((((a_1^{a_2})^{a_3})^{a_4})^{a_5}), (((a_1^{(a_2^{a_3}))^{a_4})^{a_5}), (((a_1^{a_2})^{(a_3^{a_4}))^{a_5}), ((a_1^{((a_2^{a_3})^{a_4}))^{a_5}), ((a_1^{(a_2^{(a_3^{a_4}))^{a_5}}}),$$
$$(((a_1^{a_2})^{a_3}))^{(a_4^{a_5})}), ((a_1^{(a_2^{a_3}))})^{(a_4^{a_5})}), ((a_1^{a_2})^{((a_3^{a_4})^{a_5})}), ((a_1^{a_2})^{(a_3^{(a_4^{a_5}))}}), (a_1^{(((a_2^{a_3})^{a_4})^{a_5}))}),$$
$$(a_1^{((a_2^{(a_3^{a_4}))^{a_5}))}}), (a_1^{((a_2^{a_3})^{(a_4^{a_5}))}}), (a_1^{(a_2^{((a_3^{a_4})^{a_5}))}}), (a_1^{(a_2^{(a_3^{(a_4^{a_5})})})})\}$$

Written out in their normal forms, we have

$$N^4 = \{a_2a_3a_4a_5,\ a_4a_5a_2^{a_3},\ a_2a_5a_3^{a_4},\ a_5a_2^{a_3^{a_4}},\ a_5a_2^{a_3a_4},$$
$$a_2^{a_3}a_4^{a_5},\ a_2a_3a_4^{a_5},\ a_2a_3^{a_4^{a_5}},\ a_2a_3^{a_4a_5},\ a_2^{a_3a_4a_5},$$
$$a_2^{a_5a_3^{a_4}},\ a_2^{a_3a_4^{a_5}},\ a_2^{a_3^{a_4a_5}},\ a_2^{a_3^{a_4^{a_5}}}\}$$

wherein the normal forms are from the parenthesized forms in the same order. Notice that all of these are distinct functions, but we must be cautious not to pick numbers such that these functions are equal. Let $a_1 = \frac{1}{2}$, $a_1 = \frac{1}{2^2}$, $a_1 = \frac{1}{2^3}$, $a_1 = \frac{1}{2^4}$, and $a_1 = \frac{1}{2^5}$. Plugging in these values and evaluating them gives

$$N^4 = \{2^{-14},\ 2^{-9}2^{-\frac{1}{4}},\ 2^{-7}2^{-\frac{3}{16}},\ 2^{-5}2^{-\frac{2}{2^{\frac{3}{16}}}},\ 2^{-5}2^{-\frac{1}{64}},$$
$$2^{-\frac{3}{8}}, 2^{-5}2^{-\frac{1}{8}},\ 2^{-2}2^{-\frac{3}{2^{\frac{1}{8}}}},\ 2^{-2}2^{-\frac{3}{512}},\ 2^{-\frac{1}{2048}},$$
$$2^{-\frac{1}{(16)2^{\frac{3}{16}}}},\ 2^{-\frac{1}{(4)2^{\frac{1}{8}}}},\ 2^{-\frac{2}{2^{\frac{3}{512}}}},\ 2^{-2^{(1-\frac{3}{2^{\frac{1}{8}}})}}\}.$$

To show that these are all different values, we need only approximate to three digits, which gives us

$$N^4 = \{0.000, 0.001, 0.006, 0.009, 0.030,$$
$$0.771, 0.029, 0.037, 0.248, 0.999,$$
$$0.962, 0.853, 0.251, 0.813\}.$$

Therefore they all give distinct numbers, and thus are different functions.

It is suspected that for most selections of $a_1, \ldots a_{n+1}$ between 0 and 1 that all the normal forms of an exponentiation n-expression are unique, however the general proof has been elusive. We have shown that this holds true for up to $n = 4$ via examples.

There is still room for investigation, however. We pose additional questions that still require analysis:

1. Find values $a_1, \ldots, a_{n+1}$ such that all the elements of N^n are distinct functions.
2. Show that with $a_k = \frac{1}{2^k}$ that the C_n functions of N^n are distinct.

Bibliography

[1] John Stillwell, *Mathematics and Its History*, Springer-Verlag, 15 Fifth Avenue, New York, NY 10010, 2002.

[2] P.J. Larcombe, *The 18th Century Chinese Discovery of the Catalan Numbers*, Mathematical Spectrum **32** (1999), 5–7.

[3] Richard P. Stanley, *Catalan Addendum*, 2013, `http://www-math.mit.edu/~rstan/ec/catadd.pdf`.

[4] Winfried Geyer, *On Tamari Lattices*, Discrete Mathematics **133** (1994), 99–122.