Math 331 Fall 2026, Lecture 4: Generalized associativity, cancellation and powers

website: `https://www.cip.ifi.lmu.de/~grinberg/t/26fa/`

1.4. Generalized associativity

So we are wondering if a "multi-factor product" of the form $a_1 * a_2 * \cdots * a_n$ in a semigroup has a well-defined value independent of its parenthesization. We know that this is true for $n = 3$ (by associativity) and for $n = 4$ (we just checked it); it is also true for $n = 2$ (trivially) and for $n = 1$ (even more trivially: it is just a_1 in this case). But how can we generalize this to higher n's?

The first question is how to even formalize the claim: what **is** a "parenthesization"? We could answer this using the language of planar binary trees, but let me instead give a simpler formalization. Instead of defining parenthesizations, we will define the "product" $a_1 * a_2 * \cdots * a_n$ using one specific parenthesization: namely, the parenthesization

$$a_1 * (a_2 * (a_3 * (\cdots * a_n)))$$

that corresponds to starting with a_n and then iteratively multiplying further and further factors from the left. This is very easy to define recursively. Then, after it is defined, we will show that it equals

$$(a_1 * a_2 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n)$$

for any $k \in \{1, 2, \ldots, n-1\}$; this means that its top-level parentheses can be moved to an arbitrary place (which means that their exact position does not matter, and therefore, of course, the position of parentheses at the lower levels does not matter either).

Here is our definition of $a_1 * a_2 * \cdots * a_n$ in its recursive form:

Definition 1.4.1. Let $(S, *)$ be a semigroup. We define a "product" $a_1 * a_2 * \cdots * a_n$ for any n elements $a_1, a_2, \ldots, a_n$ of S (where n is any positive integer) by recursion on n:

- For $n = 1$, we just define it to be a_1.
- For $n > 1$, we define

$$a_1 * a_2 * \cdots * a_n := a_1 * \underbrace{(a_2 * a_3 * \cdots * a_n)}_{\substack{\text{this is defined by the} \\ \text{induction hypothesis}}}.$$

If $(S, *)$ is a monoid (i.e., has a neutral element e), then we also define the "product" $a_1 * a_2 * \cdots * a_n$ for $n = 0$ by setting it to be e.

As we said, this recursive definition means that we define $a_1 * a_2 * \cdots * a_n$ as

$$a_1 * (a_2 * (a_3 * (\cdots * a_n))) .$$

Note that for $n = 2$, our newly defined product $a_1 * a_2 * \cdots * a_n$ from Definition 1.4.1 is precisely the original $a_1 * a_2$, so we are not creating a clash of notation by overriding the existing notation.

Now, the claim that "multi-factor products" do not depend on their parenthesization takes the following form (which is convenient both for proving it and for using it):

Theorem 1.4.2 (general associativity)**.** Let $(S, *)$ be a semigroup.

(a) For any integers $0 < k < n$ and any n elements $a_1, a_2, \ldots, a_n$ of S, we have

$$a_1 * a_2 * \cdots * a_n = (a_1 * a_2 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n) .$$

(b) Assume that $(S, *)$ is a monoid (i.e., has a neutral element e). Then, for any integers $0 \leq k \leq n$ and any n elements $a_1, a_2, \ldots, a_n$ of S, we have

$$a_1 * a_2 * \cdots * a_n = (a_1 * a_2 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n) .$$

Proof. **(a)** We proceed by strong induction[1] on n.

Base case: You don't need a base case when you do strong induction. But if you want, the $n = 1$ case is vacuously true, since there is no $0 < k < 1$. And the $n = 2$ case is obvious.

Induction step: Fix a positive integer n. Assume that Theorem 1.4.2 **(a)** is true for all positive integers smaller than n instead of n (that is, roughly speaking, any "multi-factor product" of fewer than n elements of S can already be reparenthesized at will). Now let us prove it for our n. So we fix an integer $0 < k < n$ and n elements $a_1, a_2, \ldots, a_n \in S$, and we aim to show that

$$\begin{aligned} & a_1 * a_2 * \cdots * a_n \\ & \stackrel{?}{=} (a_1 * a_2 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n) . \end{aligned} \tag{1}$$

From $0 < k < n$, we obtain $n > 1$, and so Definition 1.4.1 gives us

$$a_1 * a_2 * \cdots * a_n = a_1 * (a_2 * a_3 * \cdots * a_n) . \tag{2}$$

Now, if $k = 1$, then this is precisely our claim (1) (since in this case, $a_1 * a_2 * \cdots * a_k = a_1$ and $a_{k+1} * a_{k+2} * \cdots * a_n = a_2 * a_3 * \cdots * a_n$), so we are done in this case.

[1] Regular induction would work just as well here, but I just want to show an example of strong induction :)

So let us WLOG assume that $k > 1$. (The word "WLOG" means "without loss of generality".) Since $k > 1$, we have

$$a_1 * a_2 * \cdots * a_k = a_1 * (a_2 * a_3 * \cdots * a_k) \tag{3}$$

by Definition 1.4.1. Using (2) and (3), we can rewrite the equality (1) (which we must prove) as

$$\begin{aligned} &a_1 * (a_2 * a_3 * \cdots * a_n) \\ &\stackrel{?}{=} (a_1 * (a_2 * a_3 * \cdots * a_k)) * (a_{k+1} * a_{k+2} * \cdots * a_n) . \end{aligned} \tag{4}$$

So our goal is now to prove (4). But our induction hypothesis yields

$$\begin{aligned} &a_2 * a_3 * \cdots * a_n \\ &= (a_2 * a_3 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n) \end{aligned}$$

(since the left-hand side is a "multi-factor product" with $n - 1 < n$ factors). Thus, the equality (4) rewrites as

$$\begin{aligned} &a_1 * ((a_2 * a_3 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n)) \\ &\stackrel{?}{=} (a_1 * (a_2 * a_3 * \cdots * a_k)) * (a_{k+1} * a_{k+2} * \cdots * a_n) , \end{aligned} \tag{5}$$

which immediately follows from the regular associativity of $*$ (which we assumed in the definition of a semigroup), applied to the three elements a_1, $a_2 * a_3 * \cdots * a_k$ and $a_{k+1} * a_{k+2} * \cdots * a_n$. Thus, (4) is proved. Hence, (1) follows, so our induction step is complete, and we are done proving Theorem 1.4.2 **(a)**.

(b) Let $0 \leq k \leq n$ be integers, and let $a_1, a_2, \ldots, a_n$ be n elements of S. We must prove the equality

$$\begin{aligned} &a_1 * a_2 * \cdots * a_n \\ &\stackrel{?}{=} (a_1 * a_2 * \cdots * a_k) * (a_{k+1} * a_{k+2} * \cdots * a_n) . \end{aligned} \tag{6}$$

If $0 < k < n$, then this equality follows from part **(a)**; thus, we only need to handle the cases $k = 0$ and $k = n$. But this is easy: For $k = 0$, the desired equality (6) says that

$$a_1 * a_2 * \cdots * a_n = e * (a_1 * a_2 * \cdots * a_n) ,$$

whereas for $k = n$, it says that

$$a_1 * a_2 * \cdots * a_n = (a_1 * a_2 * \cdots * a_n) * e;$$

both of these follow from the neutrality of e. Thus, part **(b)** is proved. □

Thanks to Theorem 1.4.2, we can now write "products" of multiple elements of a semigroup without parentheses. We shall freely do this henceforth. Thus, some of the proofs in Lecture 2 can be simplified: for example, one of the computations in the proof of Theorem 1.3.5 **(b)** simplifies to

$$b^{-1} * \underbrace{a^{-1} * a}_{=e} * b = b^{-1} * \underbrace{e * b}_{=b} = b^{-1} * b = e.$$

1.5. Multiplicative notation

There is a shorthand notation for semigroups:

Definition 1.5.1. Let $(S, *)$ be a semigroup. Sometimes, we will write the operation $*$ as $\cdot$, so that its values $a * b$ will be written as $a \cdot b$. We can also (optionally) omit the $\cdot$ sign, so these values will simply become ab. Moreover, the neutral element e of S (if it exists, i.e., if $(S, *)$ is a monoid) will be written as 1. This is called **multiplicative notation**, or **writing** S **multiplicatively**.

When is this allowed? Whenever there is no other multiplication operation defined on the set S that would make this notation ambiguous. So, for example, we cannot write the monoid $(\mathbb{Z}, +)$ multiplicatively, since there is already a $\cdot$ operation on $\mathbb{Z}$ and we do not want $2 \cdot 3$ to suddenly be 5 (nor do we want 23 to be 5). A good example of a monoid for which multiplicative notation is legitimate and perfectly reasonable is the monoid $\operatorname{Map}(A, A)$ of maps from a given set A to itself. In fact, mathematicians often use multiplicative notation for composition of maps (i.e., they write fg for $f \circ g$), even when these maps are not part of a monoid. However, in some situations, this is not a good idea; functions on $\mathbb{R}$ or $\mathbb{C}$ are often multiplied pointwise, so fg means the function that sends each x to $f(x)\, g(x)$ instead of the composition $f \circ g$. When using notation like this, always look out for what is standard in the specific field.

The main advantages of multiplicative notation are its brevity ("ab" is shorter than "$a * b$") and its familiarity (we have muscle memory for manipulating products, some of which can be reused for "products" in arbitrary semigroups; we just need to be careful not to assume too much based on analogy with numbers).

In particular, when using multiplicative notation in a monoid S, the neutrality of its neutral element e rewrites as "$1a = a1 = a$ for each $a \in S$", which looks very natural.

When using multiplicative notation for a semigroup $(S, *)$, we will refer to the operation $*$ as **multiplication** and to its values as **products**, even if they have nothing to do with products of numbers.

1.6. Cancellation and solving equations

As you should have noticed by now, working in a monoid is much like working with (products of) square matrices; and working in a group is correspondingly like working with invertible matrices. You cannot swap factors in a product (unless you have a specific justification for that), but you can invert and move things from one side to the other:

Theorem 1.6.1. Let S be a monoid, written multiplicatively.

Let $a, b, c \in S$ be three elements. Assume that a has an inverse. (If S is a group, then this assumption is automatically satisfied.) Then:

(a) We have $ab = c$ if and only if $b = a^{-1}c$.

(b) We have $ba = c$ if and only if $b = ca^{-1}$.

(c) We have $ab = ac$ if and only if $b = c$.

(d) We have $ba = ca$ if and only if $b = c$.

Proof. **(a)** $\Longrightarrow$: If $ab = c$, then $a^{-1} \underbrace{c}_{=ab} = \underbrace{a^{-1}a}_{=1} b = 1b = b$, so that $b = a^{-1}c$.

$\Longleftarrow$: If $b = a^{-1}c$, then $ab = \underbrace{aa^{-1}}_{=1} c = 1c = c$.

(b) is analogous to **(a)** (in a very precise sense: you can obtain a proof of part **(b)** by copying the above proof of part **(a)** but reading each product right-to-left).

(c) $\Longrightarrow$: If $ab = ac$, then we obtain $b = c$ by comparing the two equalities $\underbrace{a^{-1}a}_{=1} b = b$ and $a^{-1} \underbrace{ab}_{=ac} = \underbrace{a^{-1}a}_{=1} c = c$.

$\Longleftarrow$: If $b = c$, then $ab = ac$ obviously follows.

(d) is analogous to **(c)** (again by reading all products right-to-left). □

Theorem 1.6.1 codifies some important techniques that can be used when computing in a group (and even in a monoid, as long as the appropriate elements have inverses):

- Parts **(a)** and **(b)** let you solve equations of the form $ax = c$ and $xa = c$ in a group (where x is the unknown and a and c are knowns). As a consequence, you can solve equations of the form $axb = c$ in a group for the unknown x (the solution is $x = a^{-1}cb^{-1}$).
- Parts **(c)** and **(d)** let you cancel equal factors from equalities in a group, but only if they either appear on the very left end of both sides or on the very right end of both sides. Specifically, part **(c)** lets you cancel the factor a from $ab = ac$, while part **(d)** lets you cancel it from $ba = ca$. But you cannot generally argue that $ab = ca$ entails $b = c$.

Corollary 1.6.2. Let S be a monoid, written multiplicatively. Let $a \in S$ be an element that has an inverse. Then, the maps

$$\begin{aligned} L_a : S &\to S, \\ x &\mapsto ax \end{aligned}$$

and

$$\begin{aligned} R_a : S &\to S, \\ x &\mapsto xa \end{aligned}$$

are bijective. (These maps are known as **left multiplication by** a and **right multiplication by** a, respectively; this explains their names.)

Proof. We shall show that the map L_a is bijective by checking that it is both injective and surjective:

1. Injectivity of L_a means that $ax = ay$ implies $x = y$. But this follows from Theorem 1.6.1 **(c)** (applied to $b = x$ and $c = y$).

2. Surjectivity of L_a means that each $s \in S$ can be written as $s = ax$ for some $x \in S$. But this is true, since you can take $x = a^{-1}s$, and then Theorem 1.6.1 **(a)** yields $ax = s$.

So the map L_a is both injective and surjective, thus bijective. Alternatively, you can just show that $L_{a^{-1}}$ is an inverse to L_a, and so the map L_a is invertible, i.e., bijective.

Either way, we have shown that L_a is bijective. Similarly, we can prove the same for R_a. Its inverse is $R_{a^{-1}}$. □

Let us next formalize the notion of a multiplication table (we have already seen a couple in Lecture 2):

Definition 1.6.3. Let S be a semigroup, written multiplicatively. Its **multiplication table** is a square table whose rows are indexed by the elements of S, whose columns are indexed by the elements of S, and whose (x, y)-th entry (i.e., its entry in row x and column y) is the "product" xy. (If the set S is infinite, then this is an infinite table, so we cannot write it down, but mathematically it is still a meaningful object.)

Corollary 1.6.4. Let G be a group. Then, all the maps L_a and R_a (for $a \in G$) are bijective. In other words, in the multiplication table of a group G, each row contains each element of G exactly once, and each column contains each element of G exactly once (like in a sudoku).

Proof. The "Then" claim follows from Corollary 1.6.2 (applied to $S = G$), since each $a \in G$ has an inverse. The "In other words" claim is just a restatement of the "Then" claim (since the a-th row of the multiplication table of G is a list of all values of L_a, so that it contains each element of G exactly once if and only if L_a is bijective; and the same holds for columns and R_a). □

Let us use this to determine all groups of size 3:

Example 1.6.5. Let G be a group of size 3. Thus, we can write G as $G = \{1, f, g\}$ where 1 is the neutral element of G (we write G multiplicatively).

What can its multiplication table look like? We claim that it must necessarily look as follows:

xy	$y=1$	$y=f$	$y=g$
$x=1$	1	f	g
$x=f$	f	g	1
$x=g$	g	1	f

.

We have filled in these entries step by step: First, since 1 is neutral, we must have $1 \cdot 1 = 1$ and $1f = f$ and $1g = g$ and $f1 = f$ and $g1 = g$. Then, fg cannot be f (since Corollary 1.6.4 shows that the f-row of the multiplication table contains each element of G exactly once, so it cannot contain two f's) and cannot be g (since Corollary 1.6.4 shows that the g-column of the multiplication table contains each element of G exactly once, so it cannot contain two g's), so it must be 1. That is, $fg = 1$. Then, ff must be g (since Corollary 1.6.4 shows that the f-row of the multiplication table contains each element of G exactly once, but it already contains 1 and f), and similar reasoning shows $gf = 1$ and $gg = f$.

This means that the operation ("multiplication") of a group of size 3 is uniquely determined by knowing which of its elements is neutral. In more informal language, this says that "there is only one group of size 3"... of course, after you have convinced yourself that there is such a group in the first place, i.e., that the above multiplication table really does satisfy the axioms of a group. Later we will see why there is a group of size n for every positive integer n, so this will become trivial; but for now, we can just check it by hand (associativity is the only part that requires a bit of work).

What about groups with size 4 or size 5 or size 6 ? Some foreshadowing: there is more than one group of size 4, but only one of size 5 (upon relabelling).

1.7. Powers

We have defined the product of n elements of a semigroup. If these n elements are all equal to a single element a, this product is called a power, just as for numbers. Again, just as for numbers, we can extend this definition to $n = 0$ when there is a neutral element, and to negative n when a is invertible:

Definition 1.7.1. Let S be a semigroup (written multiplicatively), and let $a \in S$ be arbitrary.

(a) For any positive integer n, we define

$$a^n := \underbrace{aa \cdots a}_{n \text{ times}}.$$

(Thus, $a^1 = a$ and $a^2 = aa$ and $a^3 = aaa$ and so on.)

(b) If S is a monoid, then we define

$$a^0 := 1 \qquad \text{(the neutral element of } S\text{)}.$$

Thus, in this case, a^n is defined for all $n \in \mathbb{N}$.

(c) If S is a monoid and a has an inverse, then we define

$$a^n := \left(a^{-1}\right)^{-n} \qquad \text{for all negative integers } n.$$

Thus, in this case, a^n is defined for all $n \in \mathbb{Z}$.

We refer to a^n as the **n-th power** of a.

What rules do you expect the powers to satisfy? For powers of numbers, we know the laws of exponents:

- We have $a^n a^m = a^{n+m}$.
- We have $(a^n)^m = a^{nm}$.
- We have $a^{-n} = \left(a^{-1}\right)^n$ (when a has an inverse).
- We have $(ab)^n = a^n b^n$.

We will soon see which of these laws still hold for elements of a semigroup (potentially assuming that an inverse exists). Not all of them do! Can you tell which ones don't?