THE BURNSIDE PROBLEM

ESHAN BAJWA

ABSTRACT. This paper gives an overview of the Burnside problem, its history, and the mechanisms of notable solutions. We explore three solutions to some of the variations of the Burnside problem, each completely separate from the other two. Two of these address the general Burnside problem, and the third covers a few small cases of the open Burnside problem.

1. INTRODUCTION

The Burnside problem is one of the oldest questions asked in group theory. An innocent, natural question that turned out to have a complex solution. In 1902, William Burnside wrote: “A still undecided point in the theory of discontinuous groups is whether the order of a group may not be finite while the order of every operation it contains is finite.” (Burnside, 1902). This was the birth of the general Burnside problem:

> **The general Burnside problem.** If a group is finitely generated, and each element is of finite order, is the group necessarily finite?

This was followed by a more specific variation that Burnside thought would be “easier”:

> **The bounded Burnside problem.** If a group G is finitely generated, and there exists some n such that $g^n = e$ for all $g \in G$, must G be finite?

And finally the open problem on free Burnside groups, which encompasses the bounded Burnside problem:

Definition 1.1. The free Burnside group $B(r, b)$ is defined as the largest group with r generators and a bound b on the order of every element. That is, for every element g there exists some $k \leq b$ such that $g^k = e$.

> **The open Burnside problem**. For which (n, m) is the free Burnside group $B(n, m)$ finite?

Trivial cases exist, such as $B(1, m) \cong C_m$. Nontrivial solutions and insights are explored in a later section.

The Burnside problems were motivated by the simple observation that all known finitely generated infinite groups were not periodic. Burnside wrote about this in his 1902 paper[1], and answered a few of the easier cases of the open problem. This set of problems heavily influenced the development of combinatorial group theory, and are still fascinating to this day.

Date: July 11, 2022.

From 1902 to 1964, much work was done on the open Burnside problem, some by Burnside himself. In 1930, a new variation called the restricted Burnside problem. This is introduced purely for historical context, and we do not cover its solution in this paper.

> **The restricted Burnside problem**. If $B(r, n)$ is finite, is its order bounded by a number dependent only on r, n?

The solution for the restricted Burnside problem of prime n was studied by Alexei Ivanovich Kostrikin in the 1950s. Much later, in 1989, Efim Isaakovich Zelmanov answered the problem in the affirmative for arbitrary exponent. He was awarded the Fields Medal in 1994 for his work.

It was not until 1964 that first solution to the general Burnside problem was found by Evgeny Golod and Igor Shafarevich, who constructed a finitely generated periodic group with infinite elements. Later, Rostislav Grigorchuk published a paper containing the construction of a Grigorchuk group, providing another solution to the general Burnside problem. This paper explores the Golod-Shafarevich construction, Grigorchuk construction, and finally some of the cases of the open Burnside problem that were addressed by Burnside himself.

2. General Background

We start with a few definitions in order to understand the problem and the ideas behind its solutions.

Definition 2.1. Groups

A group is a set equipped with an operator such that the elements of the set abide by four rules. Let us mark our operator using $*$

1. **Closure:** For any $a, b \in G$, $a * b \in G$
2. **Identity:** There exists some $e \in G$ such that for any $g \in G$, $g * e = e * g = g$
3. **Invertability:** For every $g \in G$ there exists some g^{-1} such that $g * g^{-1} = g^{-1} * g = e$ where e is the identity element.
4. **Associativity:** The operation of the group is associative. $a*b*c = (a*b)*c = a*(b*c)$ for any $a, b, c \in G$

A group is *abelian* if the group operation is commutative.

Definition 2.2. Generators

In the study of groups, a subset is a *generating set* if any element in the group can be expressed as a product of the elements in the generating set.

It can also be thought of as starting with the generating set and appending products of elements to the set. Once no new elements can be formed, the set is closed and the entire group has been generated.

Definition 2.3. Order of Groups and Elements

Groups have order, which is simply the number of elements in the group. Elements of a group also have order, but in this case it refers to the smallest positive integer n such that $g^n = e$. A group where every element is of finite order is called a periodic group.

Definition 2.4. Conjugates of Elements

The conjugate of an element g by another element h is defined as hgh^{-1}.

Proposition 2.5. *The order of an element is the same as the order of its conjugate.*

Proof. Consider the elements g and h. Let g have order n, meaning $g^n = e$. Now consider the conjugate of g by h, $hgh{-}1$. Raising this to the power n gives $(hgh^{-1}) = \underbrace{hgh^{-1}hgh^{-1}\dots hgh^{-1}}_{n \text{ times}} = \underbrace{hg(h^{-1}h)g(h^{-1}\dots h)gh^{-1}}_{n \text{ times}} = hg^nh^{-1} = heh^{-1} = e$. 🦆

We now have all the definitions required to understand the Burnside problems. However, a little more is required to understand the constructions used to solve the Burnside problems. From here onward, set-set multiplication is done by the Cartesian set product, but instead of returning a set of ordered pairs of elements, it returns a set of products of elements.

As an example, let S, S_1, S_2 be sets, g be some element, and $*$ denote some defined multiplication for g and the elements of S, S_1, S_2. Element-set (gS) and set-set (S_1S_2) multiplication is as shown below.

$$gS = \{g * s : s \in S\}, Sg = \{s * g : s \in S\}$$

$$S_1S_2 = \{s_1 * s_2 : s_1 \in S_1, s_2 \in S_2\}$$

Definition 2.6. Cosets

A coset is the product of a subgroup N with an element $a \notin N$. aN is a "left coset" and Na a "right coset". If the group is abelian, its left cosets and right cosets are the same. We call a a representative of the coset.

Proposition 2.7. *Every element in a coset can act as a representative of it.*

Proof. Let $A = a_1N$ be a left coset of G created by some $a_1 \notin N$. Let a_2 be another element in A. There must be some $b \in N$ such that $a_1b = a_2$. This gives $a_1^{-1}a_1b = a_1^{-1}a_2$ so $b = a_1^{-1}a_2$. $bN = N$ because N is a group, so $a_1^{-1}a_2N = N$. Thus $a_2N = a_1N$, and both a_1 and a_2 are representatives of A. The same concept can be applied to right cosets. 🦆

A subgroup is said to be normal if and only if it partitions the group into cosets that form a group under the set-set product defined earlier. This group is called a quotient group, and is denoted G/N for a normal subgroup N of a group G.

In order for the cosets determined by a subgroup N to form a group, closure must be shown (as well as associativity, identity, and invertability, but those are trivial). This means the product of two cosets, however it is defined, must be another coset. $AN = aNN = aN = A$ because $NN = N$ because N is closed. In order for the cosets to form a group, aNb must be another left coset of N. Let us find one element of this resulting coset by using e from N. $aeb = ab$. Therefore the resulting coset of AB must be abN. This can be further reduced to $gNg^{-1} = N$.

If $aNb = abN$ then $Nb = bN$, so left and right cosets by a normal subgroup are identical. This leads us to $bb^{-1}N = bNb^{-1} = N$. In fact, this identity is equivalent because we can also go from $bNb^{-1} = N$ to $aNb = abN$ by $aNb = a(bNb^{-1})b = abN$. We can now define the normal subgroup.

Definition 2.8. Normal Subgroups

A normal subgroup is a subgroup N that is invariant under conjugation, meaning $gNg^{-1} = N$. Notice that for an abelian group, every subgroup is normal.

3. Background for Golod-Shafarevich

The first solution to the general Burnside problem requires further knowledge of abstract algebra. While groups are powerful structures, they are limited to one operation. To model sets with multiple operations, other group-like structures with multiple operations are used. These structures are used in the Golod-Shafarevich construction.

Definition 3.1. Rings

A ring is a set of elements and two operations called addition and multiplication. These operations can be defined in any way that satisfies the conditions of a ring.

(1) **Addition**: The elements of the ring must form an abelian group under the defined addition of the ring. The additive identity is often denoted as "0".
(2) **Multiplication**: The elements of the ring must be closed under multiplication. If there is a multiplicative identity (not required), it is usually denoted as "1".
(3) **Associativity**: Multiplication must be associative.
(4) **Distribution**: The distributive property must hold. $(a)(b+c) = ab + ac$.

An interesting result of rings is that the additive identity 0 always ends up being an absorbing element by multiplication. $0a = 0$. This is easily derived using the distributive property. An element g if a ring is said to be *nilpotent* if there exists some n such that $g^n = 0$.

Rings have a property called *characteristic*, which is the smallest positive integer n such that

$$\forall a \in R, \underbrace{a + a + \cdots + a}_{\text{n times}} = 0$$

If there is a multiplicative identity in R, this definition is equivalent to

$$\underbrace{1 + 1 + \cdots + 1}_{\text{n times}} = 0$$

.

If there is no such n, the characteristic is said to be 0.

An ideal of a ring is akin to a normal subgroup of a group.

Definition 3.2. An ideal I of a ring R is a subset of R that divides the ring into cosets that themselves form a ring, called a quotient ring. This requires the following properties:

(1) I is closed under addition and multiplication
(2) I is absorbing under multiplication, meaning $\forall a \in R, aI = I$

The idea of cosets and representatives exists in rings as well. However, in rings, addition is used to form the cosets (ex. $A = a + I$ for some ideal I and representative a). Ideals are "sided" similar to cosets.

With this comes the notion of generating an ideal. Often, the set given for generation does not actually generate an ideal in the same way generators of a group generate the group. Generating an ideal actually means taking the smallest ideal of a ring that includes the elements in the given generating set.

A *field* is similar to a ring, but has inverses and identity for multiplication, and is commutative.

Definition 3.3. Fields

A field is a set of elements with two operations called addition and multiplication. These operations can be defined in any way that satisfies the conditions of a field.

(1) **Addition**: The elements of the field must form an abelian group under the defined addition of the ring. The additive identity is commonly denoted as "0".
(2) **Multiplication**: The elements of the field must form an abelian group under the defined multiplication of the ring. The multiplicative identity is usually denoted as "1".
(3) **Distribution**: The distributive property must hold. $(a)(b + c) = ab + ac$.
(4) **Scalar Commutativity**: $ax_1 * bx_2 = abx_1x_2$.

Field also have characteristics. Because fields are more restrictive than rings, some properties arise.

Proposition 3.4. *The characteristic of a field is always prime.*

Proof. Assume the characteristic k is not prime. Then k can be written as a product of two integers, α and β. Thus

$$\underbrace{\underbrace{(1+1+\cdots+1)}_{\alpha \text{ times}} + \underbrace{(1+1+\cdots+1)}_{\alpha \text{ times}} + \cdots + \underbrace{(1+1+\cdots+1)}_{\alpha \text{ times}}}_{\beta \text{ times}} = 0$$

By closure,

$$\underbrace{(1+1+\cdots+1)}_{\alpha \text{ times}} = g \in F$$

Now there are two cases.
The first case: $g = 0$. In this case,

$$\underbrace{1+1+\cdots+1}_{\alpha \text{ times}} = 0$$

$\alpha < k$ thus the field must be in characteristic α, not k.
The second case: $g \neq 0$. In this case,

$$\underbrace{(g+g+\cdots+g)}_{\beta \text{ times}} = 0$$

Consider

$$\underbrace{(g+g+\cdots+g)}_{\beta \text{ times}} * g^{-1} = 0 * g^{-1} = 0$$

By the distributive property,

$$\underbrace{(1+1+\cdots+1)}_{\beta \text{ times}} = 0$$

thus the field is in characteristic β, not k. 🦆

Proposition 3.5. *The characteristic of a field always divides the order of the field.*

This is due to Lagrange's Theorem, which states that the order of any subgroup must divide the order of the whole group. The characteristic of a field is just the order of the additive subgroup generated by 1, thus it must divide the order of the field.

Putting these two properties together, we find that if a field has a prime p elements, the characteristic of the field must be p.

An algebra over a field, or just an algebra for short, is similar to a vector space. This structure is denoted $A = F\langle x_1, x_2, \dots \rangle$ where F is a field and x_i are vectors in some space. There must be defined scalar multiplication.

For the Golod-Shafarevich construction, we describe *free* algebra.

Definition 3.6. Free algebra (over a field)

(1) **Basis**: The basis vectors of a free algebra $F\langle x_1, x_2, \dots \rangle$ are all the words formed by $\langle x_1, x_2, \dots \rangle$. These vectors are linearly independent.
(2) **Multiplication**: Multiplication of two vectors is defined as the concatenation of the words representing vectors. Multiplication is not necessarily commutative. Multiplication follows the distributive property.
(3) **Identities**: The additive identity of the vector space is the zero vector. Multiplying any vector by the additive identity 0 of the field yields the zero vector. Multiplicative identity is only present in the scalar multiplication, where $1v = v$ for any vector v.

It is also possible to take the ideal of an algebra, with the same requirements used for rings.

4. Golod-Shafarevich Construction

This construction, adapted from (uncudh, 2009), utilizes the Golod-Shafarevich theorem, which we do not prove.

Let $A = F\langle x_1, ..., x_n \rangle$ be the free algebra over a field F of non-commuting variables x_i. Let I be the two-sided ideal of A generated by homogenous elements f_j of degree d_j such that $2 \leq d_1 \leq d_2 \leq ....$ Let r_i be the number of $d_j = i$. The fundamental inequality of Golod-Shafarevich leads to:

Theorem 4.1. *If all $r_i \leq \frac{(n-1)^2}{4}$, A/I is infinite dimensional.*

Proof. This is due to (Golod & Shafarevich, 1964). 🦆

Now we show the construction of an infinite, periodic group from finite generators using a similar construction to the one shown for the theorem. Let F be a field of a prime p order. Let $T = F\langle x_1, x_2, x_3 \rangle$ be the free algebra over F. Let T' be the ideal of T consisting of all elements of T without constant term. Let T_n be the subspace of T consisting of all homogeneous elements of degree n.

List the elements of T' as $t_1, t_2, \dots$. Choose an integer $m_1 \geq 2$ and write $t_1^{m_1}$ as $v_1 + v_2 + \dots + v_{k_1}$ where $v_i \in T_i$. Now choose some large enough integer m_2 such that $t_2^{m_2}$ can be expressed as $v_{k_1+1} + v_{k_1+2} + \dots + v_{k_2}$. Continue for all $t_n \in T'$. Now let I be the smallest two-sided ideal of T containing all v_i, in other words the two-sided ideal generated by v_i.

Now let a_1, a_2, a_3 be the elements $x_1 + I, x_2 + I, x_3 + I$ of T/I. Let G be a multiplicative semi subgroup of T/I generated by the elements $1 + a_1, 1 + a_2, 1 + a_3$. G is finitely generated and we will proceed to show that it is also infinite and periodic, satisfying the conditions of the general Burnside problem.

Proposition 4.2. *G is a proper periodic group.*

Proof. t_i^m for some large enough m can be written as a sum of v_i, which is an element in I. Therefore every $a \in T'/I$ is nilpotent, as the representatives of a are elements of T', which will all land back in I after being raised to some m. For some large enough n, $a^{p^n} = 0$. We are in characteristic p due to proposition 3.5, so $(1 + a)^{p^n} = 1 + a^{p^n} = 1$. Therefore any $1 + a$ has an inverse $(1+a)^{o-1}$ where o is the smallest positive integer such that $(1+a)^o = 1$. Every element of G can be written as $1 + a$ as it was generated completely from $(1 + a_i)$, so G is a proper periodic group. 🦆

Using the Golod-Shafarevich inequality, T/I is infinite-dimensional. For our construction, $r_i = 1$ for all i, and $n = 3$. Thus we satisfy $r_i \leq \frac{(n-1)^2}{4}$ and can conclude that T/I is infinite-dimensional.

Proposition 4.3. *G is an infinite group.*

Proof. Suppose G was finite. The linear combinations of elements of G would form a finite-dimensional algebra A over F. $1+a_i$ and 1 are in G, so $(1+a_i)-1 = a_i \in A$. 1, a_1, a_2, a_3 are all in B. This set of elements generates all of T/I, which we know to be infinite-dimensional. Thus B must also be infinite-dimensional, so we have a contradiction, and G must be infinite. 🦆

We have now shown that G is an infinite, finitely generated, periodic group, answering the general Burnside problem. However, there is no bound on the order of elements, as we simply chose some large enough number m_i in our construction, so it does not answer the bounded Burnside problem.

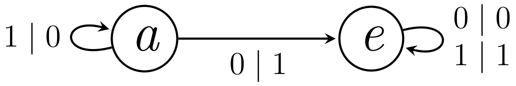
5. Automata Groups

Golod and Shaferivich's work on the burnside problem was interesting, but not direct. It was merely a byproduct of their work on class-field towers. Here, we construct an automata group that is interesting for a number of reasons. For this paper, we are only interested in its relation to the general Burnside problem. This automata require much less knowledge of advanced abstract algebra, but a general understanding of state machines. This section is adapted from (Feuilloley, 2018).

A state machine is represented as a graph. Think of using the machine as traversing the graph according to a set of instructions. The state machines used here are called mealy machines, which return an output in addition to traversing the graph. We define our input as a string of letters from some alphabet a. Think of each state as a function, which takes in a string and, depending on the first character of the string, returns the concatenation of the corresponding output and the call of the corresponding function on the rest of the string (excludes the first character). In the case that the input word is of length one, the function only returns the corresponding output, ommiting the function call.

This paper uses a standard formatting for these graphs. The labels of the nodes serve only as names and do not contribute to the output. Instead, the machine interprets the input according to the labels of the edges. Traverse the graph according to first letter (on the left of the mid line) and write to the output according to the second letter (on the right of the mid line). Once we have finished following our input "instructions" we are done.

As a simple example, we can construct a machine that adds one to any binary number. We define our alphabet $a = \{0, 1\}$ and write our words as binary numbers, but with reverse digits. For example, 8 would be written as 0001.



In this diagram, inputting a reversed binary number into state a outputs that binary number plus one in the same format. For example, we can pass in 1011 as 13. Following the graph, we can read the letter by letter path as

$$a \xrightarrow{1 \mid 0} a \xrightarrow{0 \mid 1} e \xrightarrow{1 \mid 1} e \xrightarrow{1 \mid 1} e$$

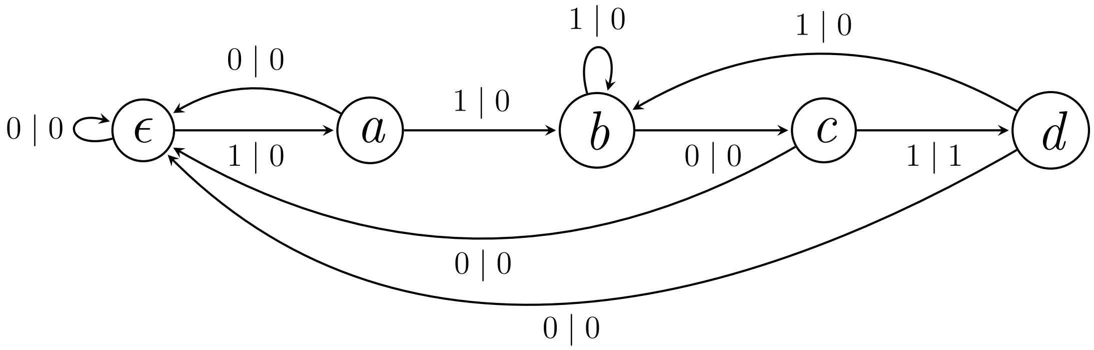
or

$$\begin{tikzcd}
 & 1 \arrow[dd] & & 0 \arrow[dd] & & 1 \arrow[dd] & & 1 \arrow[dd] & \\
a \arrow[rr] & & a \arrow[rr] & & e \arrow[rr] & & e \arrow[rr] & & e \\
 & 0 & & 1 & & 1 & & 1 &
\end{tikzcd}$$

We can also follow this path in an equation $a(1011) = 0 + a(011) = 01 + e(11) = 011 + e(1) = 0111$. Here, "+" denotes concatenation, not numerical addition. This gives us the output 0111, which is 14 in binary, reversed. Alternatively, we can tell the machine to read right to left. This way, we can input a binary number without needing to reverse it. In this case, the number 13 would be represented as 1101, and $a(1101) = a(110) + 0 = e(11) + 10 = e(1) + 110 = 1110$, which represents the number 14. Fortunately, the implementation is not very important as the abstraction of adding one still holds.

The functions on a word that correspond to inputting that word into the machine at some state can be put under function composition. From here, we check if this forms a group. e acts as the identity, since for any word w, $e(w) = w$. a is more complicated, but if we adhere to the abstraction of it adding one to the numerical value of the word, the powers of a become clear. a^n is a function that adds one to a number n times. Unfortunately, this does not generate a group because we have no way of getting an element that corresponds to subtracting a value from a number. What we can do is limit the size of our words to l digits. If we do this, then adding one to a string full of 1's will put us back at zero, because the one overflows into the place value $l + 1$. Thus $a^{2^l} = e$, generating a cyclic group of order 2^l.

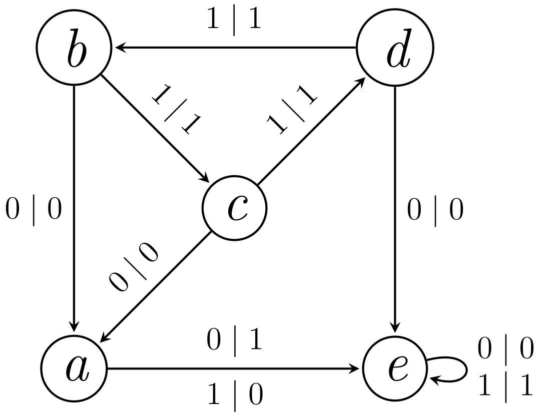
These machines have many creative use cases, especially in computer science. For example, here is a machine that detects the substring 1101 and writes a 1 to the output at the index of the last character of the substring, while the other indicies have 0 as a sort of filler element.



This works by starting at state ϵ. When the first character (1) in the substring (1101) is found, the machine changes to a state a, which looks for the second character (0). If the second character (1) appears, we move to b to check for the third character. If it finds a 0, we cannot be in the correct substring so we reset to the starting state ϵ. Note that for the state b, which checks for the third character, if we find a 1, we would have 111 and we can start looking for the substring at the second 1, meaning we want to now check for the third character, so we start at state b. This structure continues for the length of the substring, until it reaches the end and writes a 1 to the output. These machines are extremely powerful and versatile, and have much more complex and abstract applications. In this paper, we are concerned with general Burnside problem and the Grigorchuk automata.

6. The Grigorchuk Group

Now that we are familiar with automata, we can get back to the Burnside problem. This section focuses on the Grigorchuk automata and the group generated by it.

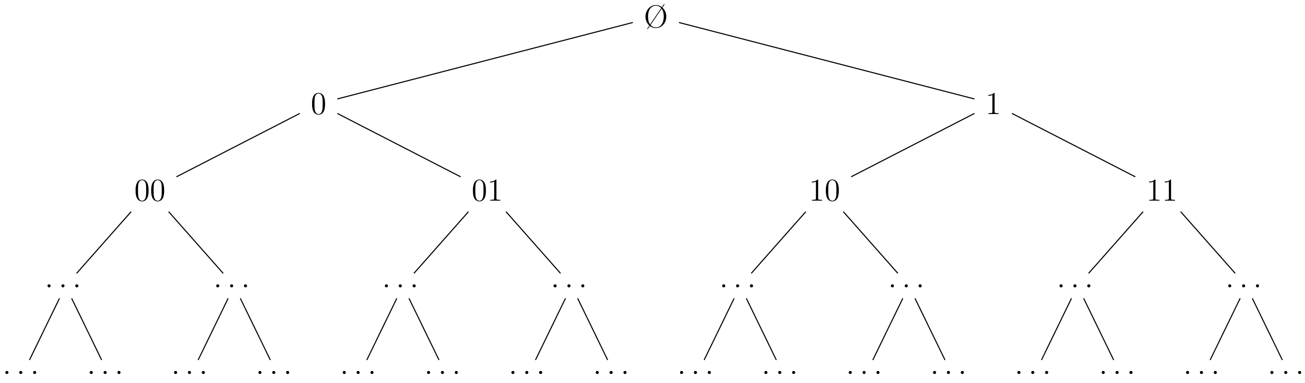


In order to construct the Grigorchuk group G, we treat the states (functions) as elements of a group under function composition, just as we did for the binary incrementer.

7. Binary Trees

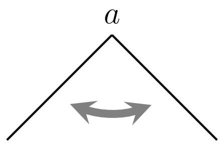
To show that this group is a solution to the general Burnside problem, let us consider these functions as automorphisms on an infinite binary tree. Instead of just taking an input and running through the machine, we represent all inputs in one tree. This section is adapted from (Hudec, 2006).

Let T_s denote the binary tree with a root node of s. The tree $T_\emptyset$ is the tree starting at the empty word and looks like this:

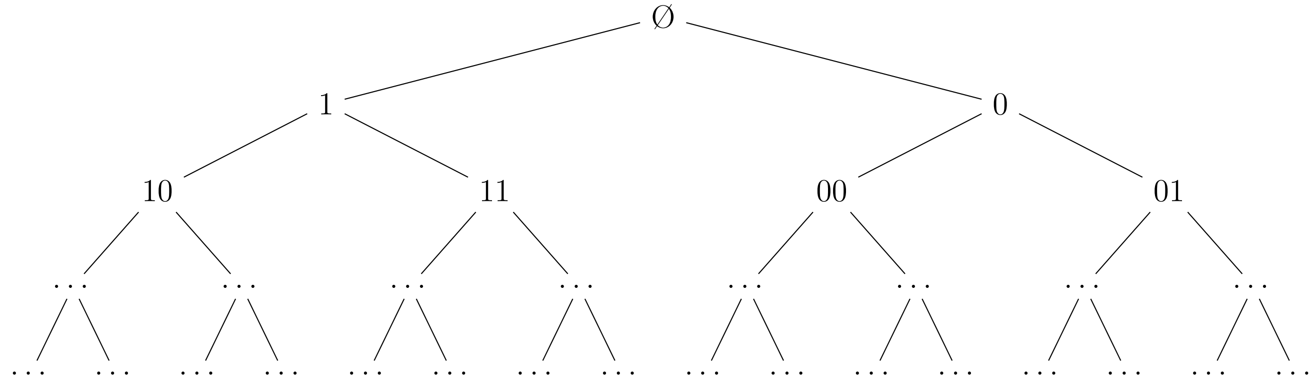


One can pick out any valid input (a word) by taking a path down the tree, defined by some sequence of left and right decisions. Each function in the Grigorchuk group G can be represented as an automorphism of the tree. The morphism, which represents an element $g \in G$, is defined such that for any word w with path p in the default binary tree, taking p in $g(T_\emptyset$ leads to the word $g(w)$. The key concept is that the tree allows us to look at the function's behavior on *all* words at the same time.

So what do the morphs look like? First off, e has no change from input to output, as it acts as a sort of identity element. a performs exactly one "flip" $(0 \mid 1,\ 1 \mid 0)$, before going to e, which means all characters after the first will not be affected. Now we can represent a as a transformation on $T_\emptyset$:



We call this a *morph tree*. This is a different structure than $a(T_\emptyset)$, which would look like this:



Notice that only the first branches are swapped. We can stop looking at the image of a branch if it goes to state e because there are no swaps in the e tree. Remember that a is a transformation, it can be applied to any tree, not just $T_\emptyset$. For example, applying a to $a(T_\emptyset)$ gives $a(a(T_\emptyset))$. This is equivalent to swapping the left and right subtrees, then swapping them again. Obviously, this leaves us back where we started, at $T_\emptyset$. This also tells us that a has an order of two.

The trees for b, c, and d are a little more complicated, as a flip (which looks like a swap in the tree) can occur at any depth into the input word, *only* when it reaches state a.

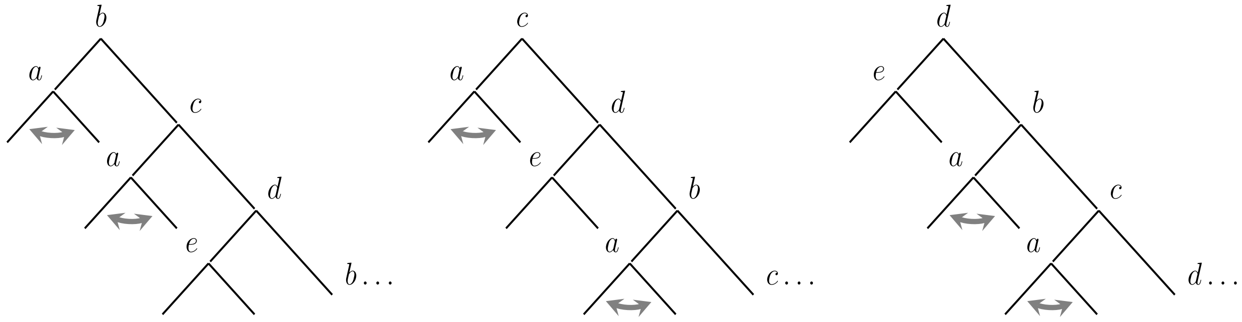


Figure 1. functions b, c, and d as automorphisms of a binary tree

Notice how b, c, and d are all roughly the same recursive structure, just starting at different states.

Think of composing functions as composing morphs to a tree. This is the same as overlaying the trees and cancelling when two swaps are in the same place. Of course, composing b with itself would yield the identity transformation, as all the swaps are in the same positions. Using this idea,

Proposition 7.1.

$$a^2 = b^2 = c^2 = d^2 = e$$

and after further observation

Proposition 7.2.

$$bc = cb = d, cd = dc = b, bd = db = c$$

Now we introduce a new way to describe transformations. Let $\phi_0(m)$ denote the left subtree of an automorphism tree m, and $\phi_1(m)$ the right subtree. Let $\psi(m) = (\phi_0(m), \phi_1(m))$. Let G be the group of automorphisms generated by a, b, c, d, e. Consider the subgroup S of G where every automorphism in S does not swap the immediate left and right subtrees of any T. For every $s \in S$, there exists a pair $(g_0, g_1) \in G \times G$ where $\psi(s) = (g_0, g_1)$. Note that ψ does not retain the full information of a transformation, as it does not capture the swap (or lack of swap) between the first two branches.

Proposition 7.3. *For two elements $s_1, s_2 \in S$, $\psi(s_1 s_2) = \psi(s_1)\psi(s_2)$. In other words, ψ of a product is the same as the product of the ψ's, if the product is in S.*

This is because the left and right subtrees remain independent when considering elements of S. As an example,

$$\psi(bc) = \psi(d) = (e, b)$$
$$\psi(b)\psi(c) = (a, c)(a, d) = (aa, cd) = (e, b)$$

ϕ_0 and ϕ_1, which are homomorphisms, have interesting properties when applied to conjugates of elements. In our group, for $\gamma \in \{b, c, d\}$, the conjugate by a is $a^{-1}\gamma a = a\gamma a$. From here on, the conjugate of an element refers to the conjugate of that element by a, unless specified.

Imagine the default binary tree T. Call the left subtree T_0 and the right subtree T_1. Apply the conjugate of γ to T and observe what it does to the tree. First, we swap T_0 and T_1 to be on the right and left, respectively. Now we apply γ. $\phi_0(\gamma)$ is applied to T_1 and $\phi_1(\gamma)$ is applied to T_0. Finally we apply a again, swapping T_0 and T_1 back to their original sides. Observe that $\phi_0(\gamma)$ was applied to the right subtree, and $\phi_1(\gamma)$ to the left. Thus $\psi(a\gamma a) = (\phi_1(\gamma), \phi_0(\gamma))$.

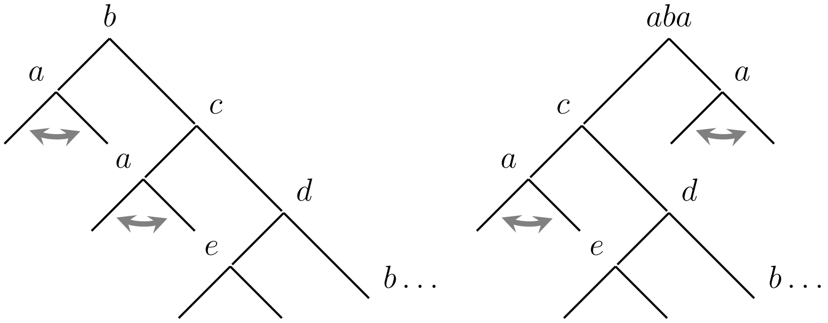


Figure 2. A comparison of b and its conjugate

Proposition 7.4. *ϕ_0 and ϕ_1 are surjective.*

Proof. Recall that G is generated by $\{a, b, c, d, e\}$, and any element can be written as a product of terms in the generating set. We will represent elements as words, where the alphabet is the generating set of the group.

$$\psi(b) = (a, c)$$
$$\psi(c) = (a, d)$$
$$\psi(d) = (e, b)$$

Applying a word to a tree T will apply some other word to both subtrees, as every element in the image of ψ is part of the generating set. In fact, it contains the entire alphabet. Using conjugates, we obtain

$$\psi(aba) = (c, a)$$
$$\psi(aca) = (d, a)$$
$$\psi(ada) = (b, e)$$

Notice how every element $\{a, b, c, d, e\}$ is in the images of both ϕ_0 and ϕ_1. Thus, we have a way of applying any word to either depth one subtree through automorphisms on the entire tree. Clearly, this extends to further depths, though they are not necessary for this proof. It naturally follows that ϕ_0 and ϕ_1 surject onto G. 🦆

We have a surjection from $S \subset G$ onto G, thus S (and consequently G) must be infinite. Another way to look at it is for any s in S, there exists some s' such that $\phi_0(s') = s$. This s' must be included in S, so for s' there is some s'', resulting in some s''', etc. All these will apply s to a deeper left subtree than the previous, thus every element generated by this process is unique. This process can also be repeated forever, so it generates an infinite number of unique elements.

Having shown the group to be infinite, and because we defined it to be finitely generated, all that is left is to show that each element is of finite order. In the Grigorchuk group, we find something more specific. The Grigorchuk group is a two-group. That is, every element has an order of a power of two.

Proposition 7.5. *The Grigorchuk group is a two-group.*

Proof. Let k be the length of the reduced word w representing an element $g \in G$ from the alphabet $\{a, b, c, d, e\}$.

Consider the bases cases $k = 0$, $k = 1$, and $k = 2$. It is not necessary to show all of them, but they are interesting. If k is 0, g must be e, the identity, thus g has an order of 1. If k is 1, $g \in \{a, b, c, d\}$ and has an order of 2.

For $k = 2$, we find that every element must be the product of a and another element, as all combinations of $\{b, c, d\}$ are reducible by proposition 7.2. Thus w is of the form $a\gamma$ or γa for any $\gamma \in \{a, b, c, d, e\}$. Observe that $(ad)^2 = (ad)(ad) = (ada)d$ which has $\psi(ada) = (b, e)$ and $\psi(d) = (e, b)$. This means $\psi(adad) = (b, b)$. b has an order of 2 and the subtrees are independent, so $adad$ has an order of 2, making ad have an order of 4. This also applied to da because we have $(da)^2 = dada = d(ada)$.

Now consider ac. Following a similar process, $(ac)^2$ has $\psi(acac) = \psi(aca)\psi(c) = (d, a)(a, d) = (da, ad)$. We know that ad and da have an order of 4, so $((ac)^2)^4$ gives $\psi(acac)^4 = (da, ad)^4 = (e, e)$. Therefore ac and ca have an order of 8.

Finally we have ab. $(ab)^2$ has $\psi(abab) = \psi(aba)\psi(b) = (c,a)(a,c) = (ca, ac)$. Both ac and ca have an order of 8, so $((ab)^2)^8$ gives $\psi(abab)^8 = (ca, ac)^4 = (e, e)$. Thus ab and ba have an order of 16.

We proceed by induction on k. Assume for all $k \geq 3$, all g of length $k-1$ or less have an order of a power of 2.

First case: k is odd. This means w must be of the form $a\gamma_1 a\gamma_2 \dots \gamma_n a$ or $\gamma_1 a\gamma_2 a \dots \gamma_m$ where $\gamma_i \in \{b, c, d\}$, because any pair of $\{b, c, d\}$ can be reduced. The first case can be written as the conjugate of some h by a. This h has at most length $k-2$, which by the induction hypothesis means h is of finite order, a power of two. By Proposition 2.5, elements have the same order as their conjugate, thus w and therefore g has finite order, a power of two. For the second case, we can take the conjugate of w by γ_1, call it $\omega = \gamma_1\gamma_1 a\gamma_2 a \dots a\gamma_m\gamma_1 = a\gamma_2 \dots a(\gamma_m\gamma_1)$. If $\gamma_m \neq \gamma_1$, then the last term $\gamma_m\gamma_1$ reduces to one of $\{b, c, d\}$, meaning ω has a length of $k-1$. If $\gamma_m = \gamma_1$, then $\gamma_m\gamma_1 = e$ and ω has reduced length $k-2$. In either case, the induction hypothesis guarantees ω has an order of a power of two. Again, any element has the same order as its conjugate, so w also has an order of a power of two. We have just shown, for odd k, g has an order of a power of 2.

Second case: k is even. We have two subcases, g is either of the from $a\gamma_1 a\gamma_2 \dots \gamma_n$ or $\gamma_1 a\gamma_2 a \dots \gamma_m a$. However, the second case has the same order as its conjugate by γ_1, written $\gamma_1 g\gamma_1$, which has the form $a\gamma_2 a\gamma_3 \dots a\gamma_1$. Thus showing the first case to have order of a power of two is sufficient to prove the same for the second case.

Again we split into two cases. Suppose the number of a's in the reduced word of g is even. Then we must have $g \in S_1$. This means we can look at $\psi(g) = \psi(a\gamma_1 a)\psi(\gamma_2)\dots\psi(\gamma_n) = (w_0, w_1)$. Notice that every individual ψ term is a pair of elements, both at most of length one. It then follows that the lengths of w_0 and w_1 are at most $\frac{k}{2}$. By the induction assumption, they both have a finite order of a power of two. g has an order of the least common multiple of the orders of w_0 and w_1, meaning it must also have order of a power of two.

Finally, we look at the last case, where k is even and the number of a's in w is odd. Consider w^2, which is twice as long as w. This is in S_1 because the number of a must be even, as the number of a's is double the number in w. Let $(\alpha, \beta) = \psi(w^2)$. Both α and β have at most a length of k, due to the same reasoning used in the previous paragraph for w_0 and w_1. We shall prove that α and β are of finite order, a power of two, which means w^2 also has finite order, again a power of two.

(1) The reduced word w^2 contains d. If it contains d, it also contains the conjugate, as if we consider w^2 as writing w twice, in the same line, we can associate the first w to have d and the second to have ada. Because $\psi(d) = (e, b)$ and $\psi(ada) = (b, e)$, the reduced words α and β have at most length $k-1$ (we can omit the e). By the induction assumption, these words have finite order, a power of two, and we are done.

(2) The reduced word w^2 contains c. Again, this means c and its conjugate aca appear somewhere in w^2. $\psi(c) = (a, d)$ and $\psi(aca) = (d, a)$. So either α and β can be reduced, and by induction we are done, or they contain d. If they contain d, we can use the same argument used in 1) to show that α^2 and β^2 (thus α and β) are of finite order, a power of two.

(3) The reduced word w^2 contains b and, consequently, its conjugate. $\psi(b) = (a, c)$, $\psi(aba) = (c, a)$. α and β, in the case that they cannot be reduced, must contain c. Then we can use the same argument used in 2) to show α^2 and β^2 (and thus α and β) are of finite order, a power of two.

All words of even length must fall into at least one of those three cases, meaning any element reducible to a word of even length must have an order of a power of two.

🦆

Finally, we have proved every element of the Grigorchuk group to be of finite order. Thus, our final condition is satisfied, and we have shown this group to be infinite, finitely generated, and periodic, answering the general Burnside problem in the negative.

8. The Open Burnside Problem

So far we have dealt with specific examples, constructions, that answered the general Burnside problem. Interestingly, though perhaps not surprisingly, approaches to the open Burnside problem were not similar. This section is adapted from (Goh, 2020).

Besides $B(1,n) \cong C_n$, there are a few more small cases, which Burnside mentioned in his 1902 paper. $B(r,2)$ is one of them. Take elements s, t. Each element has an order of two, so $s^2 = t^2 = (st)^2 = e$. Now $stst = e$ and by multiplying by s on the left and t on the right we find $st = ts$ thus $B(r,s)$ is abelian. This completely defines the quotient group $(\mathbb{Z}/2\mathbb{Z})^r$, thus the groups must be isomorphic and $B(r,2)$ has order $|(\mathbb{Z}/2\mathbb{Z})^r| = 2^r$.

Burnside groups of exponent three have also been solved.

Proposition 8.1. *For $r \geq 1$ the order of $B(r,3)$ is $3^{m(r)}$ for $m(r) \leq 3^{r-1}$.*

Proof. The group is a three group, and has an order of a power of three. Note that $(st)^3 = e$ implies

$$\begin{aligned}(s^{-1}t^{-1}s^{-1})(ststst) &= (s^{-1}t^{-1}s^{-1})e \\ &= tst = s^{-1}t^{-1}s^{-1} \qquad (1)\end{aligned}$$

for any s, t in the group.

When $r = 1$, our group is of order three, thus we can find a $m(1) \leq 3^0$ by $m(1) = 1$. Proceed by induction on r.

Assume our claim $|B(k,3)| = 3^{m(k)}$ holds for some integer k. Form $B(k+1,3)$ by adding one new generator to the set of generators. We can write any element $s \in B(k+1,3)$ as

$$s = s_1 g^{\pm 1} s_2 g^{\pm 1} \dots g^{\pm 1} s_n$$

Our goal is to now reduce the number of g's through reductions on s, mostly using the identity (1). The first reduction we make is for consecutive g's of the same sign. By (1) we have

$$g s_i g = s_i^{-1} g^{-1} s_i^{-1} \text{ and } g^{-1} s_i g^{-1} = s_i^{-1} g s_i^{-1}$$

where t is $g^{\pm 1}$. Though the length of the word is the same, this representation is more useful for us because it reduces the number of g's, thus we will write s in a form with alternating signs of g. Remember $g^3 = e$, so we can make further reductions.

$$\begin{aligned}s &= s_1 \dots s_i g s_{i+1} g^{-1} s_{i+2} g \dots s_n \\ &= s_1 \dots s_i g s_{i+1} g g s_{i+2} g \dots s_n \\ &= s_1 \dots s_i (g s_{i+1} g)(g s_{i+2} g) \dots s_n \\ &= s_1 \dots s_i (s_{i+1}^{-1} g^{-1} s_{i+1}^{-1})(s_{i+2}^{-1} g^{-1} s_{i+2}^{-1}) \dots s_n \\ &= s_1 \dots s_i s_{i+1}^{-1} (g^{-1} s_{i+1}^{-1} s_{i+2}^{-1} g^{-1}) s_{i+2}^{-1} \dots s_n\end{aligned}$$

and we can continue this reduction process over and over again. This allows us to write any element in $B(k+1,3)$ with a maximum of two $g^{\pm 1}$'s. Through further reductions, we see

$$s_1 g^{-1} s_2 g s_3 = s_1 (g^{-1} s_2 g^{-1}) g^{-1} s_3 = s_1 s_2^{-1} g s_2^{-1} g^{-1} s_3$$

so any s can be represented as one of the forms

$$s_1,\ s_1 g s_2,\ s_1 g^{-1} s_2,\ s_1 g s_2 g^{-1} s_3$$

For the first form, we have $3^{m(k)}$ possibilities as $|B(k,3)| = 3^{m(k)}$. For the second and third forms, we find all combinations of two elements, where order matters, so we get $3^{m(k)} 3^{m(k)} = 3^{2m(k)}$. For the fourth form, we have $3^{3m(k)}$. Thus

$$|B(k+1,3)| = 3^{m(k)} + 2 * 3^{2m(k)} + 3^{3m(k)} = 3^{m(k+1)} < 3^{3m(k)+1}$$

$m(r)$ is an integer so $m(k+1) \leq 3m(k) \leq 3^k$ thus we satisfy $m(r) \leq 3^{r-1}$ and $|B(k+1)| \leq 3^{3^k}$. 🦆

In 1993, F.W. Levi and B.L. van der Waerden found the exact value of $m(r)$ to be

$$m(r) = \binom{r}{3} + \binom{r}{2} + r$$

Surprisingly, Burnside groups of exponent larger than three are much more mysterious. We do not know the orders of $B(r,4)$ for $r > 5$. However, $B(r,4)$ has been shown to be finite for any r.

For n not equal to $2, 3, 4, 6$ it is not known whether or not $B(r,n)$ is finite. However, it has been shown that there are n such that $B(r,n)$ is infinite. This was produced by Novikov and Adian in 1968, when they proved $B(r,n)$ infinite for all odd $n \geq 4381$, and solved the bounded Burnside problem. This result was improved to odd $n \geq 665$ in by Adian in 1975, and then to odd $n \geq 101$ in 2015, again by Adian.

This century old problem has lead to great advancements in many areas of abstract algebra, and still remains partly unsolved to this day. Even something as small as $B(2,5)$ has unknown order.

References

Burnside, W. (1902). On an unsettled question in the theory of discontinuous groups. *Quart. J. Pure and Appl. Math.*, *33*, 230–238.

Feuilloley, L. (2018, Jun). *Introduction to mealy automata, grigorchuk group and the burnside problem.* Retrieved from `https://semidoc.github.io/godin-mealy`

Goh, M. K. (2020, Apr). *The burnside problem.* Retrieved from `https://marcelgoh.ca/misc/expo/burnside.pdf`

Golod, E. S., & Shafarevich, I. R. (1964). On the class field tower. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, *28*(2), 261–272.

Hudec, D. A. (2006). *On the burnside problem.* Retrieved from `https://math.uchicago.edu/~may/VIGRE/VIGRE2006/PAPERS/Hudec.pdf`

uncudh. (2009, Nov). *The general burnside problem.* Retrieved from `https://uniformlyatrandom.wordpress.com/2009/09/04/the-general-burnside-problem/`