REU 2005 · Discrete Mathematics · Lecture 6

Instructor: László Babai
Scribes: T. Schedler and Murali Ganapathy

June 27, 2005. Last updated June 30, 2005

1 Lecture 6

1.1 Presentations (in terms of Generators and Relations)

A *prsentation* of a commutative monoid is a description of the monoid as the largest monoid satisfying a certain set of equations (relations) among a set of generators.

Since we will be working in commutative monoids, we will use additive notation.

Let $\mathcal{M} = \langle a, b \mid 5a = 7b \rangle$, be unique largest monoid generated by a and b subject to the relation $5a = 7b$. What is $\mathcal{M}$? Consider the following monoid $\mathcal{N}$ which satisfies the above relations: $\mathcal{N} = \langle 7, 5 \rangle \subseteq (\mathbb{N}, +) = \{7x + 5y \mid x, y, \in \mathbb{N}\}$. By the definition of $\mathcal{M}$, we have a **epimorphism** (i.e. a surjective homomorphism) $\mathcal{M} \twoheadrightarrow \mathcal{N}$.

Since $\mathcal{M}$ is generated by a and b, and satisfies $5a = 7b$, every element of $\mathcal{M}$ can be written as $x \cdot a + y \cdot b$, where $0 \leq x \leq 4$ and $y \in \mathbb{N}$. Now consider the homomorphism φ from $\mathcal{M} \to \mathcal{N}$ which maps $xa + yb \in \mathcal{M}$ to $7x + 5y \in \mathcal{N}$. By choice of $\mathcal{M}$, φ is surjective.

Now every element of $\mathcal{N}$ can be written as $7x + 5y$ for a unique choice of $0 \leq x \leq 4$ and $0 \leq y$. Now suppose $\varphi(x_1a + y_1b) = \varphi(x_2a + y_2b)$. Then we have $7x_1 + 5y_1 = 7x_2 + 5y_2$. This implies $7x_1 = 7x_2 \pmod 5$. Since $\gcd(5, 7) = 1$, it follows that $x_1 = x_2 \pmod 5$ and hence $x_1 = x_2$ (since $0 \leq x_1, x_2 \leq 4$). Now $x_1 = x_2$ implies $y_1 = y_2$.

This shows that the map φ is also injective (i.e. one-to-one). Hence φ is an isomorphism, and $\mathcal{M} \cong \mathcal{N}$.

The upshot is that there are multiple ways to describe the same monoid. In our particular example, we have a "canonical form" for the minimal presentation, i.e. $\mathcal{N}$. In general finding such a "canonical form" is a very difficult problem. Actually, one can show that it is undecidable whether two presentations give rise to the same (noncommutative) group or not or even whether a group given by a presentation is just the identity. So in general it may be impossible to find a canonical form of the elements of a monoid.

1.2 The sandpile monoid

Recall: The "ambient space" is a digraph. It has "ordinary vertices" and a "sink". $V = V_0 \cup \{sink\}$. A configuration f is a map $f : V_0 \to \mathbb{N}$. It is **stable** if $(\forall v \in V_0)(f(v) \leq \deg^+(v) - 1)$.

So $\mathcal{M}$ is the set of stable configurations under $\oplus$. Here $f \oplus g := \sigma(f+g)$ where σ is the stabilization operator (keep firing until stable).

Exercise 1.1. $|\mathcal{M}| = \prod_{v \in V_0} \deg^+(v)$.

Elements of the minimal ideal $\mathcal{G} \triangleleft \mathcal{M}$ are called **recurrent configurations**. Recall $\mathcal{M} = \langle t_v : v \in V_0 \rangle$, where t_v has just a single chip, which is at vertex v.

Observation 1.2. f is recurrent iff $\forall g \in \mathcal{M}$, there exists $h \in \mathcal{M}$ such that $f = g \oplus h$, i.e. f is reachable from any configuration g.

Proof. $\Rightarrow$: Take any $g \in \mathcal{M}$ and consider the ideal $g \oplus \mathcal{M} \triangleleft \mathcal{M}$ generated by g. Minimality of $\mathcal{G}$ implies $\mathcal{G} \subseteq g \oplus \mathcal{M}$. Hence for any $f \in \mathcal{G}$, we have $f \in g \oplus \mathcal{M} = \{g + h \mid h \in \mathcal{M}\}$.

$\Leftarrow$: Suppose f is reachable from any configuration. We need to show that $f \in \mathcal{G}$. Let $g \in \mathcal{G}$ be arbitrary. Since f is reachable from g, if follows that $f = g \oplus h$ for some appropriate $h \in \mathcal{M}$. Then we have that $f \in \mathcal{G} \oplus h \subseteq \mathcal{G}$ because $\mathcal{G}$ is an ideal. So $f \in \mathcal{G}$. ☐

Now $|\mathcal{G}|$ is the number of recurrent configurations, which is the number of spanning trees oriented toward the sink (by a theorem from last time we didn't prove).

Definition 1.3. A **functional subgraph** H is a subgraph such that $\deg^+_H(v) = 1$ for all $v \in V_0$ and $\deg^+_H(\text{sink}) = 0$.

Equivalently functional subgraphs are in one-to-one correspondence (if there are no parallel edges), to functions $f : V_0 \to V$ such that $\forall v \in V_0, (v, f(v))$ is an edge of the ambient space.

Exercise 1.4. *The number of functional subgraphs is equal to* $\prod \deg^+(v)$ *(the choices for* $f(v)$ *for different* v *are independent).*

Corollary 1.5. $|\mathcal{M}|$ = *number of functional subgraphs.*

Now, spanning trees directed towards the sink are the same as functional subgraphs H without a cycle. These are called **acyclic** functional subgraphs.

Corollary 1.6. —*G*— = *number of acyclic functional subgraphs.*

Recall that $\mathcal{T} = \{\text{transient configurations}\} = \mathcal{M} \setminus \mathcal{G}$.

Corollary 1.7. —$\mathcal{T}$— = *number of functional subgraphs containing a cycle.*

1.3 The "big cycle" example

We studied an example, which is a directed cycle of $(n-1)$ vertices, where each vertex is connected to the sink by seven additional edges. So $|\mathcal{M}| = 8^{n-1}$ where $|V| = n$ and $|V_0| = n-1$. There is only one cyclic functional subgraph namely the cycle on V_0, and hence only one transient state. Hence $|\mathcal{T}| = 1$. This element of $\mathcal{T}$ has to be 0 because of the previous result (if $0 \notin \mathcal{T}$ then $\mathcal{T} = \emptyset$). Hence $|\mathcal{G}| = 8^{n-1} - 1$.

Lemma 1.8. *("blank cycles lemma") Let C be a directed cycle in the ambient space. If w is a configuration with ≥ 1 chip on C, then the same is true for all configurations reachable from w.*

Proof. Suppose not. Then starting from w there is a sequence of chip additions and topplings, which results in emptying out all chips on C. Considering the configuration just before C becomes empty, we have a stable configuration w' reachable from w, and a vertex v such that w' has at least one chip on C and $w' \oplus t_v$ does not have any chip on C.

Let vertex u be the last vertex in C to be toppled. Then the vertex following it on C will end up non-empty.

$\square$

Claim 1.9. *In our example, $\mathcal{G}$ is cyclic.*

Proof. We see that $8t_v = t_{v+1}$ where $v+1$ is the vertex after v;

It follows that $t_2 = 8t_1$, $t_3 = 8t_2 = 64t_1$, $\cdots, t_i = 8^{i-1}t_1$. Since $\mathcal{G} = \langle t_1, \cdots, t_{n-1}\rangle$, it follows that $\mathcal{G} = \langle t_1 \rangle$ is cyclic. $\square$

Definition 1.10. The configuration $w_{\max}$ is the one in which every vertex v has $\deg^+(v) - 1$ chips. This is called the "saturated configuration."

Since $w_{\max}$ is reachable from any stable configuration, it follows that

Lemma 1.11. *$w_{max} \in \mathcal{G}$.*

Claim 1.12. *In our example, w_{max} is the identity of $\mathcal{G}$.*

Proof. Since $t_v \in \mathcal{G}$ for every $v \in V_0$ (since $\mathcal{G} = \mathcal{M} \setminus \{0\}$), the identity e must satisfy $t_v \oplus e = t_v$. If $e(v) < 7$ then $t_v + e$ is a stable configuration and hence cannot be equal to t_v. Thus we must have $e(v) = 7$, i.e. the identity element of $\mathcal{G}$ is the saturated. $\square$

In general we know that the configuration where every vertex is saturated is recurrent (it is reachable from all stable configurations). Although this is the identity in our example, it's not always the identity of $\mathcal{G}$ ($t_v \in \mathcal{G}$ need not hold for a general ambient space).

For example, in the 200 × 200-grid case, we got a beautiful picture for the identity shown in the last class.

1.4 When is $\mathcal{G} = \mathcal{M}$?

Now let's consider the question: when is $\mathcal{G} = \mathcal{M}$

Claim 1.13. *0 is recurrent iff every state is recurrent, i.e. $0 \in \mathcal{G}$ iff $\mathcal{G} = \mathcal{M}$.*

Proof. $\Leftarrow$: clear. $\Rightarrow$: if $0 \in \mathcal{G}$, then $w \in \mathcal{M}$ implies that $w = 0 \oplus w$ is reachable from 0. 0 is reccurent implies it is reachable from all configurations. But all configurations are reachable from 0. Hence all configurations are reachable from each other, i.e. $\mathcal{G} = \mathcal{M}$. $\square$

We have the following charecterization:

Claim 1.14. *$0 \in \mathcal{G}$ iff the ambient space is a DAG (directed acyclic graph).*

Proof. $\Rightarrow$: Suppose the ambient space has a cycle C. The $w_{\max}$ configuration is always in $\mathcal{G}$ and has atleast one chip on C. From the cycle lemma, all configurations reachable from $w_{\max}$ must have at least one chip on C. But $0 \in \mathcal{G}$ implies 0 is reachable from $w_{\max}$, a contradiction.

$\Leftarrow$: (Proof 1: draining process) Since the ambient space is a DAG, it is possible to assign numbers $1, 2, \ldots, n$ to each vertex, in such a way, that all edges in the ambient space go to the right, i.e. a low numbered vertex to a high numbered vertex. Let w be any configuration, and look at the lowest numbered vertex which has a chip on it. Now add chips to that vertex till it fires. After it fires, there are no chips on that vertex and more importantly on all vertices lower than it. Hence the number of empty vertices has gone up by one. Eventually, all vertices will be empty, i.e. the 0 configuration is reachable from every stable configuration.

(Proof 2: Counting) $|\mathcal{M}|$ = the number of functional subgraphs and $|\mathcal{G}|$ = the number of acyclic functional subgraphs. Also $\mathcal{G} \subseteq \mathcal{M}$. If the ambient space is a DAG, all functional subgraphs are acyclic. Hence $|\mathcal{M}| = |\mathcal{G}|$. Hence $\mathcal{G} = \mathcal{M}$. ☐

1.5 The structure of finitely generated abelian groups

Recall from group theory the definition of **direct product** of groups: $G = G_1 \times G_2 \times \cdots \times G_k = \{(g_1, \ldots g_k) : g_i \in G_i\}$ with the multiplication $(g_1, \ldots, g_k)(g'_1, \ldots, g'_k) = (g_1 g'_1, \ldots, g_k g'_k)$.

Exercise 1.15. *$C_k \times C_\ell$ is cyclic iff $\gcd(k, \ell) = \ell$.*

Now if G_i are abelian groups and $G = \prod_{i=1}^{k} G_i$, then $G_1 \cong \{(g_1, 1_2, 1_3, \ldots, 1_k) : g_1 \in G_1\}$. Denote this latter subgroup of G by $\widetilde{G}_1$. We have $G_i \cong \widetilde{G}_i \leq G$. We also note that

1. $\widetilde{G}_1 \times \cdots \times \widetilde{G}_k = G$; and
2. $\widetilde{G}_i \cap \prod_{j \neq i} \widetilde{G}_1 = \{\mathrm{id} = (1_1, 1_2, \ldots, 1_k)\}$.

We have an internal characterization of the direct product: if G is an abelian group, and $H_1, \ldots, H_k \leq G$, then G is a direct product of the H_i ($G \cong H_1 \times \cdots \times H_k$) iff

1. $H_1 \cdots H_k = G$; and
2. $(\forall i) H_i \cap \prod_{j \neq i} H_j = \{1\}$.

Now we have the **Fundamental theorem of finitely generated abelian groups**. What does it mean to be finitely generated? That there is a finite set of elements which generate the group. What groups are not that way? An important example is $(\mathbb{Q}, +)$. (Another example: the free abelian group on infinitely many generators. But not everyone needs to know what this is.) The theorem goes as follows:

Theorem 1.16. *If G is a finitely generated abelian group, then G is the direct product of a finite number of cyclic groups.*

Moreover, we can write $G = C_{n_1} \times C_{n_1} \times \cdots \times C_{n_k}$ where $2 \leq n_1 \mid n_2 \mid n_3 \mid \ldots \mid n_k$ (where we substitute 0 for infinity: recall $n \mid 0$ for any number n). This sequence of invariants n_k is unique:

Exercise 1.17. *Prove the sequence of invariants (n_i) are unique. (First, assume that G is finite).*

We have the

Definition 1.18. The rank of G is k.

Exercise 1.19. *The rank is the minimal number of generators.*

As a special case, $\mathbb{Z} \times \cdots \times \mathbb{Z} = \mathbb{Z}^k$ cannot be generated by fewer than k generators ("dimension invariance").

We have the following result due to Dhar:

Exercise 1.20. *(hard): The rank of the sandpile group of the $n \times n$-grid is n.*

Open question: are there infinitely many values of n such that the sandpile group of the $n \times (n+1)$-grid is cyclic?

We can consider modular reduction: if G is an abelian group (written additively), we can consider $G/2G$, which is a quotient satisfying $2x = 0$. We have the

Exercise 1.21. $\mathit{rank}(G) \geq \mathit{rank}(G/2G)$.

We get that $G/2G \cong C_2 \times \cdots \times C_2$, $\operatorname{rank}(G/2G)$ times. Call this $\operatorname{rk}_2(G) := \operatorname{rk}(G/2G)$, or the **2-rank** of G.

Exercise 1.22. $\operatorname{rk}_2(n \times m \text{ \textit{grid}}) = \gcd(n+1, m+1) - 1$. *In particular,* $\operatorname{rk}_2(n \times n \text{ \textit{grid} }) = n$.

Exercise 1.23. $\min(n, m) \geq \operatorname{rk}(n \times m \text{ \textit{grid}}) \geq \gcd(n+1, m+1) - 1$.