0.1. A generalized principle of inclusion/exclusion

Exercise 1. Let \( n \in \mathbb{N} \). Let \( S \) be a finite set. Let \( A_1, A_2, \ldots, A_n \) be finite subsets of \( S \). Let \( k \in \mathbb{N} \). Let \( S_k \) be the set of all elements of \( S \) that belong to exactly \( k \) of the subsets \( A_1, A_2, \ldots, A_n \). (In other words, let \( S_k = \{ s \in S \mid \text{the number of } i \in [n] \text{ satisfying } s \in A_i \text{ equals } k \} \).) Prove that

\[
|S_k| = \sum_{I \subseteq [n]} (-1)^{|I|-k} \binom{|I|}{k} \left| \bigcap_{i \in I} A_i \right|
\]

Here, the intersection \( \bigcap_{i \in \emptyset} A_i \) is understood to mean the whole set \( S \).

Note that the principle of inclusion and exclusion (see, e.g., [Galvin17, §16]) is the particular case of Exercise 1 for \( k = 0 \) (since \( S_0 = S \setminus \bigcup_{i=1}^{n} A_i \)).

Exercise 1 is a result of Charles Jordan (see [Comtet74, §4.8, Theorem A] and [DanRot78] for fairly complicated proofs). I further generalize it in [Grinbe16, Theorem 2.39] (make sure to download the version of December 25th!). Let me here give a self-contained proof.

First, we recall two facts from the solutions to homework set #5:

Proposition 0.1. We have

\[
\binom{m}{n} = 0
\]

for every \( m \in \mathbb{N} \) and \( n \in \mathbb{N} \) satisfying \( m < n \).

Corollary 0.2. Let \( n \in \mathbb{N} \). Let \( i \in \mathbb{N} \). Then,

\[
\sum_{j=0}^{n} (-1)^{j} \binom{n}{j} \binom{j}{i} = (-1)^i [n = i].
\]
Next, we recall the classical formula for the size of a subset using Iverson brackets:

**Lemma 0.3.** Let $S$ be a finite set. Let $T$ be a subset of $S$. Then,

$$|T| = \sum_{s \in S} [s \in T].$$

Lemma 0.3 allows us to reduce a formula for $|S_k|$ to a formula for $[s \in S_k]$ (for any given $s \in S$). Here is the latter formula:

**Lemma 0.4.** Let $n \in \mathbb{N}$. Let $S$ be a finite set. Let $A_1, A_2, \ldots, A_n$ be finite subsets of $S$. Let $S_k$ be the set of all elements of $S$ that belong to exactly $k$ of the subsets $A_1, A_2, \ldots, A_n$. (In other words, let $S_k = \{s \in S \mid \text{the number of } i \in [n] \text{ satisfying } s \in A_i \text{ equals } k\}$.) Let $s \in S$. Then,

$$[s \in S_k] = \sum_{I \subseteq [n]} (-1)^{|I| - k} \binom{|I|}{k} \left[ s \in \bigcap_{i \in I} A_i \right].$$

Here, the intersection $\bigcap_{i \in \emptyset} A_i$ is understood to mean the whole set $S$.

**Proof of Lemma 0.4** Define a subset $C$ of $[n]$ by

$$C = \{i \in [n] \mid s \in A_i\}.$$

Thus, for each $i \in [n]$, we have the following equivalence:

$$(i \in C) \iff (s \in A_i). \quad (1)$$

But recall the definition of $S_k$. From this definition, we obtain the following equivalence:

$$(s \in S_k) \iff \left( \text{the number of } i \in [n] \text{ satisfying } s \in A_i \text{ equals } k \right) = |\{i \in [n] \mid s \in A_i\}| \iff \left( \left\{ i \in [n] \mid s \in A_i \right\} \text{ equals } k \right) = |C| \iff \left( |C| = k \right).$$

Hence, we find the following equality between truth values:

$$[s \in S_k] = [|C| = k]. \quad (2)$$
On the other hand, let $I$ be any subset of $[n]$. Then, we have the following equivalence:

\[
\left( s \in \bigcap_{i \in I} A_i \right) \iff \left( \begin{array}{l}
s \in A_i \\
\text{for each } i \in I \\
\iff (i \in C) \\
\text{(by (1))}
\end{array} \right) \iff (i \in C \text{ for each } i \in I) \\
\iff (I \subseteq C).
\]

Thus,

\[
\left[ s \in \bigcap_{i \in I} A_i \right] = [I \subseteq C]. \quad (3)
\]

Now, forget that we fixed $I$. We thus have proven (3) for each subset $I$ of $[n]$. 
Thus,

\[\sum_{I \subseteq [n]} (-1)^{|I|-k} \binom{|I|}{k} \left[ \sum_{s \in \bigcap_{i \in I} A_i} \right] = [I \subseteq C] \text{(by 3)}\]

\[= \sum_{I \subseteq [n]} (-1)^{|I|-k} \binom{|I|}{k} [I \subseteq C]\]

\[= \sum_{I \subseteq C; \ I \subseteq [n]} (-1)^{|I|-k} \binom{|I|}{k} + \sum_{I \subseteq [n]; \ not \ I \subseteq C} (-1)^{|I|-k} \binom{|I|}{k} = 0\]

\[= \sum_{j \in \mathbb{N}} \sum_{|I|=j} (-1)^{|I|-k} \binom{|I|}{k}\]

\[= (-1)^{j-k} \binom{j}{k}\]

\[= (\text{the number of all subsets } I \text{ of } C \text{ satisfying } |I|=j) (-1)^{j-k} \binom{j}{k}\]

\[= \sum_{j \in \mathbb{N}} \binom{|C|}{j} (-1)^{j-k} \binom{j}{k}\]

\[= \sum_{j \in \mathbb{N}} (-1)^{j-k} \binom{|C|}{j} \binom{j}{k}\]
\[
\sum_{j=0}^{\infty} (-1)^j \binom{|C|}{j} \binom{j}{k} + \sum_{j=|C|+1}^{\infty} (-1)^{j-k} \binom{|C|}{j} \binom{j}{k}
\]

(by Proposition 0.1, applied to \(|C|\) and \(j\) instead of \(m\) and \(n\) (since \(|C| < j\) (because \(j \geq |C| + 1 > |C|\)))

(here, we have split the summation at \(j = |C|\))

\[
= \sum_{j=0}^{|C|} (-1)^j (-1)^k \binom{|C|}{j} \binom{j}{k} + \sum_{j=|C|+1}^{\infty} (-1)^{j-k} 0 \binom{j}{k}
\]

\[
= \sum_{j=0}^{|C|} (-1)^j (-1)^k \binom{|C|}{j} \binom{j}{k} = (-1)^k \sum_{j=0}^{|C|} (-1)^j \binom{|C|}{j} \binom{j}{k}
\]

(by Corollary 0.2, applied to \(|C|\) and \(k\) instead of \(n\) and \(i\))

\[
= (-1)^k (-1)^k [|C| = k] = [|C| = k] = [s \in S_k] \quad (by \ 2).
\]

This proves Lemma 0.4.

\[\square\]

\textbf{Solution to Exercise 1}  

For every subset \(I\) of \([n]\), the intersection \(\bigcap_{i \in I} A_i\) is a subset of \(S\).

Thus, for every subset \(I\) of \([n]\), we have

\[
\left| \bigcap_{i \in I} A_i \right| = \sum_{s \in S} \left[ s \in \bigcap_{i \in I} A_i \right]
\]

\[\quad (by \ 1)\]

\[\quad (In \ fact, \ this \ is \ obvious \ when \ I \ is \ nonempty \ (because \ all \ the \ A_i \ are \ subsets \ of \ S), \ but \ it \ also \ holds \ when \ I \ is \ empty \ (because \ in \ this \ case, \ the \ intersection \ \bigcap_{i \in I} A_i = \bigcap_{i \in \emptyset} A_i \ is \ defined \ to \ be \ S).)\]
(by Lemma 0.3, applied to $T = \bigcap_{i \in I} A_i$). Hence,

$$\sum_{I \subseteq [n]} (-1)^{|I| - k} \binom{|I|}{k} \left| \bigcap_{i \in I} A_i \right| = \sum_{s \in S} \left| s \cap \bigcap_{i \in I} A_i \right|$$

$$= \sum_{I \subseteq [n]} (-1)^{|I| - k} \binom{|I|}{k} \sum_{s \in S} \left| \bigcap_{i \in I} A_i \right| = \sum_{I \subseteq [n]} \sum_{s \in S} (-1)^{|I| - k} \binom{|I|}{k} \left| s \cap \bigcap_{i \in I} A_i \right|$$

$$= \sum_{s \in S} \sum_{I \subseteq [n]} (-1)^{|I| - k} \binom{|I|}{k} \left| s \cap \bigcap_{i \in I} A_i \right| = \sum_{s \in S} [s \in S_k].$$

Comparing this with $|S_k| = \sum_{s \in S_k} [s \in S_k]$ (which follows from Lemma 0.3 applied to $T = S_k$), we obtain

$$|S_k| = \sum_{I \subseteq [n]} (-1)^{|I| - k} \binom{|I|}{k} \left| \bigcap_{i \in I} A_i \right|. $$

This solves Exercise 1.  

0.2. Summing fixed point numbers of permutations

Recall that for any $n \in \mathbb{N}$, we let $S_n$ denote the set of all permutations of $[n]$. If $S$ is a finite set, and if $f : S \to S$ is a map, then we let $\text{Fix } f$ denote the set of all fixed points of $f$. (That is, $\text{Fix } f = \{s \in S \mid f(s) = s\}$.)

Exercise 2. Let $n$ be a positive integer. Prove that $\sum_{w \in S_n} |\text{Fix } w| = n!$.

[Hint: Rewrite $|\text{Fix } w|$ as $\sum_{i \in [n]} [w(i) = i].$]

(In other words, this exercise states that the average number of fixed points of a permutation of $[n]$ is 1.)

Exercise 2 was Problem 1 at the International Mathematical Olympiad (IMO) 1987.

Our solution to Exercise 2 relies on the following facts:

Lemma 0.5. Let $m \in \mathbb{N}$. Let $G$ be an $m$-element set. Then, the number of all permutations of $G$ is $m!$. 

Proof of Lemma 0.5 (sketched). There is a bijection \( \alpha : G \to [m] \) (since \( G \) is an \( m \)-element set). Fix such an \( \alpha \). Then, the map
\[
\{ \text{permutations of } G \} \to \{ \text{permutations of } [m] \},
\sigma \mapsto \alpha \circ \sigma \circ \alpha^{-1}
\]
is also a bijection\(^2\). Hence,
\[
|\{ \text{permutations of } G \}| = |\{ \text{permutations of } [m] \}|
= (\text{the number of all permutations of } [m]) = m!.
\]
In other words, the number of all permutations of \( G \) is \( m! \). This proves Lemma 0.5. \( \square \)

Lemma 0.6. Let \( n \) be a positive integer. Let \( i \in [n] \). Then, the number of all permutations \( w \in S_n \) satisfying \( w(i) = i \) is \( (n - 1)! \).

Proof of Lemma 0.6 (sketched). Roughly speaking, a permutation \( w \in S_n \) satisfying \( w(i) = i \) is “nothing but” a permutation of the \( (n - 1) \)-element set \([n] \setminus \{i\}\) (because it has to map \( i \) to \( i \), and therefore must map the remaining elements of \([n]\) to elements other than \( i \)). This is not rigorous, because strictly speaking a permutation of \([n]\) cannot be a permutation of \([n] \setminus \{i\}\) (after all, the former has domain \([n]\) while the latter only has domain \([n] \setminus \{i\}\)). Here is a rigorous version of the above statement:

To each permutation \( w \in S_n \) satisfying \( w(i) = i \), we can assign a permutation \( \tilde{w} \) of \([n] \setminus \{i\}\) by letting
\[
\tilde{w}(p) = w(p) \quad \text{for each } p \in [n] \setminus \{i\}.
\]
This defines a map
\[
A : \{ w \in S_n \mid w(i) = i \} \to \{ \text{permutations of } [n] \setminus \{i\} \},
\quad w \mapsto \tilde{w}.
\]
(4)

Conversely, to each permutation \( u \) of \([n] \setminus \{i\}\), we can assign a permutation \( \hat{u} \in S_n \) satisfying \( \hat{u}(i) = i \) by setting
\[
\hat{u}(p) = \begin{cases} u(p), & \text{if } p \neq i; \\ i, & \text{if } p = i \end{cases} \quad \text{for each } p \in [n].
\]
This defines a map
\[
B : \{ \text{permutations of } [n] \setminus \{i\} \} \to \{ w \in S_n \mid w(i) = i \},
\quad u \mapsto \hat{u}.
\]

\(^2\)This is straightforward to verify.
The maps \( A \) and \( B \) are well-defined and mutually inverse\(^5\). Thus, there is a bijection from the set \( \{ w \in S_n \mid w(i) = i \} \) to the set \( \{ \text{permutations of } [n] \setminus \{i\} \} \) (namely, \( A \)). Hence,

\[
|\{ w \in S_n \mid w(i) = i \}| = |\{ \text{permutations of } [n] \setminus \{i\} \}|
= (\text{the number of all permutations of } [n] \setminus \{i\})
= (n - 1)!
\]

(by Lemma \( 0.5 \) (applied to \( G = [n] \setminus \{i\} \) and \( m = n - 1 \)), because \( [n] \setminus \{i\} \) is an \((n - 1)\)-element set). In other words, the number of all permutations \( w \in S_n \) satisfying \( w(i) = i \) is \((n - 1)! \). This proves Lemma \( 0.6 \). \( \square \)

**Solution to Exercise 2** (sketched). If \( w \in S_n \) and \( i \in [n] \), then

\[
| i \in \text{Fix } w | = | w(i) = i |
\]

(5)

If \( w \in S_n \), then \( \text{Fix } w \) is a subset of \([n]\), and therefore Lemma \( 0.3 \) (applied to \( S = [n] \) and \( T = \text{Fix } w \)) yields

\[
| \text{Fix } w | = \sum_{s \in [n]} | s \in \text{Fix } w | = \sum_{i \in [n]} | \underbrace{i \in \text{Fix } w}_{\text{by (5)}} | [w(i) = i] \quad \left( \text{here, we have renamed the} \right.
\]

\[
= \sum_{i \in [n]} | w(i) = i |.
\]

(6)

But if \( i \in [n] \), then \( \{ w \in S_n \mid w(i) = i \} \) is a subset of \( S_n \), and therefore Lemma \( 0.3 \) (applied to \( S = S_n \) and \( T = \{ w \in S_n \mid w(i) = i \} \)) yields

\[
| \{ w \in S_n \mid w(i) = i \} | = \sum_{s \in S_n} \left[ s \in \{ w \in S_n \mid w(i) = i \} \right] \left( \iff (s(i)=i) \right)
= \sum_{s \in S_n} | s(i) = i | = \sum_{w \in S_n} | w(i) = i |
\]

(7)

(here, we have renamed the summation index \( s \) as \( w \)).

---

\(^5\)This is straightforward to check (just remember that permutations are bijective).

\(^4\)because of the equivalence \((i \in \text{Fix } w) \iff (i \text{ is a fixed point of } w) \iff (w(i) = i) \)
Now,
\[
\sum_{w \in S_n} \frac{\text{wt}(w)}{\text{Fix}(w)} = \sum_{w \in S_n} \sum_{i \in [n]} \frac{[w(i) = i]}{[w(i) = i]} = \sum_{i \in [n]} \sum_{w \in S_n} \frac{[w(i) = i]}{[w = w(i)]}
\]
(by (6))
\[
= \sum_{i \in [n]} \sum_{w \in S_n} \frac{[\{w \in S_n \mid w(i) = i\}]}{[n]}
\]
(= the number of all permutations \(w \in S_n\) satisfying \(w(i) = i\))
(by Lemma 0.6)
\[
= \sum_{i \in [n - 1]} (n - 1)! = \frac{[n - 1]! \cdot (n - 1)!}{n!} = n \cdot (n - 1)! = n!.
\]

This solves Exercise 2. \(\square\)

**Remark 0.7.** Exercise 2 can be generalized: If \(n \in \mathbb{N}\) and \(k \in \mathbb{N}\) satisfy \(n \geq k\), then
\[
\sum_{w \in S_n} \binom{\text{wt}(w)}{k} = (n - k)! \binom{n}{k} = \frac{n!}{k!}.
\]
Do you see how the above solution can be extended to cover this generalization?

### 0.3. Transpositions \(t_{1,i}\) generate permutations

Recall a basic notation regarding permutations:

**Definition 0.8.** Let \(n \in \mathbb{N}\). Let \(i\) and \(j\) be two distinct elements of \([n]\). We let \(t_{i,j}\) be the permutation in \(S_n\) which switches \(i\) with \(j\) while leaving all other elements of \([n]\) unchanged. Such a permutation is called a transposition.

**Exercise 3.** Let \(n \in \mathbb{N}\). Prove that each permutation in \(S_n\) can be written as a composition of some of the transpositions \(t_{1,2}, t_{1,3}, \ldots, t_{1,n}\).

(Note that this composition can be empty – in which case it is understood to be \(\text{id}\) –, and it can contain any given transposition multiple times.)

To solve this exercise, we recall another definition:

**Definition 0.9.** Let \(n \in \mathbb{N}\). Let \(i \in [n - 1]\). Then, \(s_i\) denotes the permutation \(t_{i,i+1} \in S_n\).

We shall use the following well-known fact ([Grinbe16 Exercise 4.1 (b)]):

**Lemma 0.10.** Let \(n \in \mathbb{N}\). Each permutation in \(S_n\) can be written as a composition of some of the transpositions \(s_1, s_2, \ldots, s_{n-1}\).

The following is easy to check:
Lemma 0.11. Let \( n \in \mathbb{N} \). Let \( i \in [n - 1] \) be such that \( i > 1 \). Then, \( s_i = t_{1,i+1} \circ t_{1,j} \circ t_{1,j+1} \).

Lemma 0.11 can be proven by straightforward verification (just check how \( s_i \) and \( t_{1,i+1} \circ t_{1,j} \circ t_{1,j+1} \) transform a given element of \([n]\), depending on whether this element is 1, \( i \) or \( i + 1 \) or something else). Let us give a slightly more skillful argument. The following fact is simple and well-known ([Grinbe16, Exercise 4.16 (a)]):

Lemma 0.12. Let \( n \in \mathbb{N} \). Let \( k \in [n] \). For every \( \sigma \in S_n \) and every \( k \) distinct elements \( i_1,i_2,\ldots,i_k \) of \([n]\), we have

\[
\sigma \circ \text{cyc}_{i_1,i_2,\ldots,i_k} \circ \sigma^{-1} = \text{cyc}_{\sigma(i_1),\sigma(i_2),\ldots,\sigma(i_k)}. 
\]

Proof of Lemma 0.11 (sketched). From \( i \in [n - 1] \), we obtain \( i \leq n - 1 \). But \( i > 1 \), so that \( i \geq 2 \), and thus \( 2 \leq i \leq n - 1 \). Hence, \( n \geq 3 \). Thus, \( 2 \in [n] \).

Clearly, \( t_{u,v} = \text{cyc}_{u,v} \) (8) for any two distinct elements \( u \) and \( v \) of \([n]\). Applying this to \( u = 1 \) and \( v = i \), we obtain \( t_{1,j} = \text{cyc}_{1,j} \).

Now, let \( \sigma = t_{1,j+1} \). Thus, \( \sigma(1) = i + 1 \) and \( \sigma(i) = i \) (since \( i \) equals neither 1 nor \( i + 1 \)). But Lemma 0.12 (applied to \( k = 2, i_1 = 1 \) and \( i_2 = i \)) yields

\[
\sigma \circ \text{cyc}_{1,j} \circ \sigma^{-1} = \text{cyc}_{\sigma(1),\sigma(i)} = \text{cyc}_{i+1,i} \quad (9)
\]

(since \( \sigma(1) = i + 1 \) and \( \sigma(i) = i \)).

The permutation \( \sigma \) is a transposition (since \( \sigma = t_{1,i+1} \)), and hence an involution. In other words, \( \sigma^{-1} = \sigma \).

But the definition of \( s_i \) yields

\[
s_i = t_{i,i+1} = \text{cyc}_{i,i+1} \quad \text{(by (8))}
= \text{cyc}_{i+1,i} = \underbrace{\sigma \circ \text{cyc}_{1,j} \circ \sigma^{-1}}_{=t_{1,i+1} \circ t_{1,j} \circ t_{1,j+1}} \quad \text{(by (9))}
= t_{1,i+1} \circ t_{1,j} \circ t_{1,j+1}.
\]

This proves Lemma 0.11. \( \square \)

Solution to Exercise 3 (sketched). We first show the following fact:

Observation 1: Let \( i \in [n - 1] \). Then, \( s_i \) can be written as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \).
[Proof of Observation 1: If \( i > 1 \), then this follows immediately from Lemma \([0.11]\). Thus, for the rest of this proof, we WLOG assume that we don’t have \( i > 1 \). Hence, \( i = 1 \). Thus, \( s_i = s_1 = t_{1,2} \) (by the definition of \( s_i \)). Thus, again it is clear that \( s_i \) can be written as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \). This proves Observation 1.]

Now, let \( \sigma \in S_n \) be a permutation. We want to write \( \sigma \) as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \).

First write \( \sigma \) as a composition of some of the transpositions \( s_1, s_2, \ldots, s_{n-1} \). (This is possible according to Lemma \([0.10]\).) Next, write each of these transpositions \( s_1, s_2, \ldots, s_{n-1} \) as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \). (This is possible according to Observation 1.) The resulting expression is now a representation of \( \sigma \) as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \).

Now, forget that we fixed \( \sigma \). We thus have shown that each \( \sigma \in S_n \) has a representation as a composition of some of the transpositions \( t_{1,2}, t_{1,3}, \ldots, t_{1,n} \). This solves Exercise \([3]\) \( \Box \)

### 0.4. \( V \)-permutations as products of cycles

Recall the following notation:

**Definition 0.13.** Let \( X \) be a set. Let \( k \) be a positive integer. Let \( i_1, i_2, \ldots, i_k \) be \( k \) distinct elements of \( X \). We define \( \text{cyc}_{i_1, i_2, \ldots, i_k} \) to be the permutation of \( X \) that sends \( i_1, i_2, \ldots, i_k \) to \( i_2, i_3, \ldots, i_k, i_1 \), respectively, while leaving all other elements of \( X \) fixed. In other words, we define \( \text{cyc}_{i_1, i_2, \ldots, i_k} \) to be the permutation of \( X \) given by

\[
\text{cyc}_{i_1, i_2, \ldots, i_k} (p) = \begin{cases} 
i_{j+1}, & \text{if } p = i_j \text{ for some } j \in \{1, 2, \ldots, k\}; \\
p, & \text{otherwise}
\end{cases} 
\]

for every \( p \in X \),

where \( i_{k+1} \) means \( i_1 \).

**Exercise 4.** Let \( n \in \mathbb{N} \). For each \( r \in [n] \), let \( c_r \) denote the permutation \( \text{cyc}_{r, r-1, \ldots, 2, 1} \in S_n \). (Thus, \( c_1 = \text{cyc}_1 = \text{id} \) and \( c_2 = \text{cyc}_{2,1} = s_1 \).)

Let \( G = \{g_1 < g_2 < \cdots < g_p\} \) be a subset of \([n]\). (The notation “\( G = \{g_1 < g_2 < \cdots < g_p\} \)” is simultaneously saying that \( G = \{g_1, g_2, \ldots, g_p\} \) and that \( g_1 < g_2 < \cdots < g_p \).)

Let \( \sigma \in S_n \) be the permutation \( c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_p} \).

Prove the following:

(a) We have \( \sigma (1) > \sigma (2) > \cdots > \sigma (p) \).

(b) We have \( \sigma ([p]) = G \).

(c) We have \( \sigma (p+1) < \sigma (p+2) < \cdots < \sigma (n) \).

(Note that a chain of inequalities that involves less than two numbers is considered to be vacuously true. For example, Exercise 4(c) is vacuously true when \( p = n - 1 \) and also when \( p = n \).)
Solution to Exercise 4 (sketched). For each \( r \in [n] \) and \( i \in [n] \), we have
\[
c_r(i) = \begin{cases} 
  r, & \text{if } i = 1; \\
  i - 1, & \text{if } 1 < i \leq r; \\
  i, & \text{if } i > r 
\end{cases}
\] (by the definition of \( c_r \)). Thus, each \( r \in [n] \) satisfies
\[
c_r(2) < c_r(3) < \cdots < c_r(n)
\] (because the one-line notation of the permutation \( c_r \) is \((r, 1, 2, \ldots, r+1, r+2, \ldots, n)\), which shows immediately that \( c_r \) is strictly increasing on the set \( \{2, 3, \ldots, n\} \)).
Moreover, each \( r \in [n] \) and \( i \in [n] \) satisfy
\[
c_r(i) \geq i - 1.
\] (This is easy to check using (10).)

We have \( g_1 < g_2 < \cdots < g_p \). Define a further integer \( g_0 \) by \( g_0 = 0 \). Then, the chain of inequalities \( g_1 < g_2 < \cdots < g_p \) can be extended to \( g_0 < g_1 < \cdots < g_p \) (since each of \( g_1, g_2, \ldots, g_p \) is \( > 0 = g_0 \)).
For each \( q \in \{0, 1, \ldots, p\} \), we let \( c_q \) denote the permutation \( c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_q} \in S_n \). Thus,
\[
\sigma_0 = c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_0} = (\text{empty composition of permutations}) = \text{id}
\] and
\[
\sigma_p = c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_p} = \sigma.
\]
Notice that
\[
\sigma_q = \sigma_{q-1} \circ c_{g_q} \quad \text{for each } q \in [p]
\] \(^5\)
Now, we claim the following:

**Observation 1:** For each \( q \in \{0, \ldots, p\} \), the following holds:

(a) We have \( \sigma_q(i) = g_{q+1-i} \) for each \( i \in [q] \).

(b) We have \( \sigma_q(j) = j \) for each \( j \in [n] \) satisfying \( j > g_q \).

(c) We have \( \sigma_q(q+1) < \sigma_q(q+2) < \cdots < \sigma_q(n) \).

(d) We have \( g_q \geq q \).

\(^5\)Proof of (13): Let \( q \in [p] \). Then, the definition of \( \sigma_{q-1} \) yields \( \sigma_{q-1} = c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_{q-1}} \). But the definition of \( \sigma_q \) yields
\[
\sigma_q = c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_q} = (c_{g_1} \circ c_{g_2} \circ \cdots \circ c_{g_{q-1}}) \circ c_{g_0} = \sigma_{q-1} \circ c_{g_0} = \sigma_{q-1} \circ c_{g_q}.
\]
This proves (13).
Proof of Observation 1: We shall prove Observation 1 by induction.\textsuperscript{6}

Induction base: Let us prove Observation 1 for $q = 0$. To do so, we must prove the following four statements:

\begin{itemize}
  \item [(a_0)] We have $\sigma_0 (i) = g_{0+1-i}$ for each $i \in [0]$.
  \item [(b_0)] We have $\sigma_0 (j) = j$ for each $j \in [n]$ satisfying $j > g_0$.
  \item [(c_0)] We have $\sigma_0 (0+1) < \sigma_0 (0+2) < \cdots < \sigma_0 (n)$.
  \item [(d_0)] We have $g_0 \geq 0$.
\end{itemize}

But all of these four statements are obvious. Indeed, \((a_0)\) is vacuously true (since there exist no $i \in [0]$); furthermore, \((b_0)\) and \((c_0)\) are obvious (since $\sigma_0 = \text{id}$); finally, \((d_0)\) follows from $g_0 = 0$. Thus, Observation 1 has been proven for $q = 0$. This completes the induction base.

Induction step: Let $h \in [p]$. Assume that Observation 1 holds for $q = h - 1$. We must now prove that Observation 1 holds for $q = h$.

We have assumed that Observation 1 holds for $q = h - 1$. In other words, the following four statements hold:

\begin{itemize}
  \item [(a_1)] We have $\sigma_{h-1} (i) = g_{(h-1)+1-i}$ for each $i \in [h-1]$.
  \item [(b_1)] We have $\sigma_{h-1} (j) = j$ for each $j \in [n]$ satisfying $j > g_{h-1}$.
  \item [(c_1)] We have $\sigma_{h-1} (h) < \sigma_{h-1} (h+1) < \cdots < \sigma_{h-1} (n)$.
  \item [(d_1)] We have $g_{h-1} \geq h - 1$.
\end{itemize}

We must prove that Observation 1 holds for $q = h$. In other words, we must prove the following four statements:

\begin{itemize}
  \item [(a_2)] We have $\sigma_h (i) = g_{h+1-i}$ for each $i \in [h]$.
  \item [(b_2)] We have $\sigma_h (j) = j$ for each $j \in [n]$ satisfying $j > g_h$.
  \item [(c_2)] We have $\sigma_h (h+1) < \sigma_h (h+2) < \cdots < \sigma_h (n)$.
  \item [(d_2)] We have $g_h \geq h$.
\end{itemize}

\textsuperscript{6}It is rather important to prove the four parts of Observation 1 \textbf{together}, rather than trying to prove them separately. This way, they can “lend each other a hand” in the induction step (as we will see below).
Recall that \( g_0 < g_1 < \cdots < g_p \). Thus, \( g_{h-1} < g_h \), so that \( g_h > g_{h-1} \). Thus, \( g_h \geq g_{h-1} + 1 \) (since \( g_h \) and \( g_{h-1} \) are integers). But (d) yields \( g_{h-1} \geq h - 1 \), so that \( g_{h-1} + 1 \geq h \). Hence, \( g_h \geq g_{h-1} + 1 \geq h \). This proves statement (d).

Let \( r = g_h \). Thus, \( r \in [n] \) (since \( h \in [p] \) and thus \( g_h \in [n] \)). Applying (13) to \( q = h \), we obtain
\[
\sigma_h = \sigma_{h-1} \circ c_{g_h} = \sigma_{h-1} \circ c_r \quad \text{(since } g_h = r).\]

Statement (c) is easy to derive from statement (c) with the help of (11).
Statement (b) easily follows from statement (b) with the help of (10).

Applying (10) to \( i = 1 \), we obtain \( c_r(1) = r = g_h > g_{h-1} \). Hence, statement (b) (applied to \( j = c_r(1) \)) yields \( \sigma_{h-1}(c_r(1)) = c_r(1) = g_h \). But from \( \sigma_h = \sigma_{h-1} \circ c_r \), we obtain
\[
\sigma_h(1) = (\sigma_{h-1} \circ c_r)(1) = \sigma_{h-1}(c_r(1)) = g_h = g_{h+1-1}. \quad (14)
\]

Finally, statement (a) can be derived from statement (a) using (14).

\[\text{Proof.}\] We want to prove statement (c). In other words, we want to prove that \( \sigma_h(h+1) < \sigma_h(h+2) < \cdots < \sigma_h(n) \). In other words, we want to prove that \( \sigma_h(k) < \sigma_h(k+1) \) for each \( k \in \{h+1, h+2, \ldots, n-1\} \). So let us fix \( k \in \{h+1, h+2, \ldots, n-1\} \). We must prove \( \sigma_h(k) < \sigma_h(k+1) \).

We have \( k \in \{h+1, h+2, \ldots, n-1\} \). Thus, \( k \geq h+1 \geq 2 \) (since \( h \geq 1 \)). But (11) yields \( c_r(2) < c_r(3) < \cdots < c_r(n) \). Thus, \( c_r(k) < c_r(k+1) \) (since \( k \geq 2 \)).

Also, \( (12) \) yields \( c_r(k) \geq k-1 \geq h \) (since \( k \geq h+1 \)). Thus, \( c_r(k) \in \{h, h+1, \ldots, n\} \).

Also, \( (12) \) yields \( c_r(k+1) \geq (k+1)-1 = k \geq k-1 \geq h \). Thus, \( c_r(k+1) \in \{h, h+1, \ldots, n\} \).

Statement (c) says that the map \( \sigma_{h-1} \) is strictly increasing on the set \( \{h+1, h+2, \ldots, n\} \). In other words, if \( u \) and \( v \) are two elements of \( \{h, h+1, \ldots, n\} \) satisfying \( u < v \), then \( \sigma_{h-1}(u) < \sigma_{h-1}(v) \).

Applying this to \( u = c_r(k) \) and \( v = c_r(k+1) \), we obtain \( \sigma_{h-1}(c_r(k)) < \sigma_{h-1}(c_r(k+1)) \) (since \( c_r(k) < c_r(k+1) \)), and since both \( c_r(k) \) and \( c_r(k+1) \) are elements of \( \{h, h+1, \ldots, n\} \).

But \( \sigma_h = \sigma_{h-1} \circ c_r \), and thus
\[
\sigma_h(k) = (\sigma_{h-1} \circ c_r)(k) = \sigma_{h-1}(c_r(k)) < \sigma_{h-1}(c_r(k+1)) = (\sigma_{h-1} \circ c_r)(k+1) = \sigma_h(k+1).
\]

This completes our proof of statement (c).

\[\text{Proof.}\] Let \( j \in [n] \) be such that \( j > g_h \). We want to show that \( \sigma_h(j) = j \).

We have \( j > g_h = r \). Thus, \( (10) \) (applied to \( i = j \)) simplifies to \( c_r(j) = j \). But \( j > g_h > g_{h-1} \); therefore, statement (b) yields \( \sigma_h(j) = j \).

Now, recall that \( \sigma_h = \sigma_{h-1} \circ c_r \). Hence,
\[
\sigma_h(j) = (\sigma_{h-1} \circ c_r)(j) = \sigma_{h-1}(c_r(j)) = \sigma_{h-1}(j) = j.
\]

This proves statement (b).

\[\text{Proof.}\] Let us prove statement (a). In other words, let us prove that \( \sigma_h(i) = g_{h+1-i} \) for each \( i \in [h] \).

Indeed, let \( i \in [h] \). We must prove that \( \sigma_h(i) = g_{h+1-i} \).

If \( i = 1 \), then this follows from (14). Hence, for the rest of this proof, we WLOG assume that \( i \neq 1 \). Thus, \( i > 1 \). Combined with \( i \in [h] \), this yields \( i \in \{2, 3, \ldots, h\} \), so that \( i-1 \in [h-1] \). Therefore, statement (a) (applied to \( i-1 \) instead of \( i \)) yields \( \sigma_{h-1}(i-1) = g_{(h-1)+1-(i-1)} = g_{h+1-i} \) (since \( h-1+1-(i-1) = h+1-i \)).

But \( i \in \{2, 3, \ldots, h\} \), so that \( 1 < i \leq h \leq r \) (because \( r = g_h \geq h \)). The equality (10) simplifies to
We have now proven all four statements (a₃), (b₃), (c₃) and (d₃). Thus, Observation 1 holds for \( q = h \). This completes the induction step; thus, Observation 1 is proven.\]

Now, we can apply Observation 1 to \( q = p \). As a result, we obtain the following four statements:

(a₃) We have \( \sigma_p(i) = g_{p+1-i} \) for each \( i \in [p] \).

(b₃) We have \( \sigma_p(j) = j \) for each \( j \in [n] \) satisfying \( j > g_p \).

(c₃) We have \( \sigma_p(p+1) < \sigma_p(p+2) < \cdots < \sigma_p(n) \).

(d₃) We have \( g_p \geq p \).

Statement (c₃) says that \( \sigma_p(p+1) < \sigma_p(p+2) < \cdots < \sigma_p(n) \). In view of \( \sigma_p = \sigma \), this rewrites as \( \sigma(p+1) < \sigma(p+2) < \cdots < \sigma(n) \). This solves Exercise 4(b).

Statement (a₃) says that \( \sigma_p(i) = g_{p+1-i} \) for each \( i \in [p] \). In view of \( \sigma_p = \sigma \), this rewrites as

\[
\sigma(i) = g_{p+1-i} \quad \text{for each } i \in [p].
\]

In other words,

\[
(\sigma(1), \sigma(2), \ldots, \sigma(p)) = (g_{p}, g_{p-1}, \ldots, g_{1}).
\]

Hence,

\[
\{\sigma(1), \sigma(2), \ldots, \sigma(p)\} = \{g_{p}, g_{p-1}, \ldots, g_{1}\} = \{g_{1}, g_{2}, \ldots, g_{p}\} = G
\]

(since \( G = \{g_{1} < g_{2} < \cdots < g_{p}\} = \{g_{1}, g_{2}, \ldots, g_{p}\} \)). Hence

\[
\sigma\left(\begin{bmatrix} [p] \\ \vdots \end{bmatrix}\right)_{i=1}^{p} = \sigma(\{1, 2, \ldots, p\}) = \{\sigma(1), \sigma(2), \ldots, \sigma(p)\} = G.
\]

This solves Exercise 4(b).

Finally, recall that \( g_1 < g_2 < \cdots < g_p \). In other words, \( g_p > g_{p-1} > \cdots > g_1 \). In view of \([15]\), this rewrites as follows: \( \sigma(1) > \sigma(2) > \cdots > \sigma(p) \). This solves Exercise 4(a). \( \square \)

---

\( c_r(i) = i - 1 \) (since \( 1 < i \leq r \)). Now, recall that \( \sigma_h = \sigma_{h-1} \circ c_r \). Thus,

\[
\sigma_h(i) = (\sigma_{h-1} \circ c_r)(i) = \sigma_{h-1}\left(c_r(i)_{i=1}^{r-1}\right) = \sigma_{h-1}(i-1) = g_{h+1-i}.
\]

Thus, \( \sigma_h(i) = g_{h+1-i} \) is proven, as we wanted. This completes the proof of statement (a₂).
Permutations $\sigma \in S_n$ satisfying the inequalities $\sigma (1) > \sigma (2) > \cdots > \sigma (p)$ and $\sigma (p + 1) < \sigma (p + 2) < \cdots < \sigma (n)$ for some $p \in \{0, 1, \ldots, n\}$ are known as “V-permutations” (as their plot looks somewhat like the letter “V”: first decreasing for a while, then increasing). Can you guess how permutations $\sigma \in S_n$ satisfying $\sigma (1) < \sigma (2) < \cdots < \sigma (p)$ and $\sigma (p + 1) > \sigma (p + 2) > \cdots > \sigma (n)$ are called? \footnote{Answer: They are called “Λ-permutations”. Both names “V-permutations” and “Λ-permutations” are due to the shape of the plot when the permutation is plotted in 2D.}

Exercise 4 is a lemma in the theory of free Lie algebras (see \cite{BleLau92}, (10)).

TODO: Explain how Exercise 4 can also be obtained as a particular case of the
formula for a permutation in terms of its Rothe diagram. (See https://sumidiot.
blogspot.com/2008/05/rothe-diagram.html for now.)

0.5. Lexicographic comparison of permutations

Definition 0.14. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation. For any $i \in [n]$, we let $\ell_i (\sigma)$ denote the number of $j \in \{i + 1, i + 2, \ldots, n\}$ such that $\sigma (i) > \sigma (j)$.

For example, if $\sigma$ is the permutation of $[5]$ written in one-line notation as $[4, 1, 5, 2, 3]$, then $\ell_1 (\sigma) = 3$, $\ell_2 (\sigma) = 0$, $\ell_3 (\sigma) = 2$, $\ell_4 (\sigma) = 0$ and $\ell_5 (\sigma) = 0$.

Definition 0.15. Let $n \in \mathbb{N}$. Let $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be two $n$-tuples of integers. We say that $(a_1, a_2, \ldots, a_n) <_{\text{lex}} (b_1, b_2, \ldots, b_n)$ if and only if there exists some $k \in [n]$ such that $a_k \neq b_k$, and the smallest such $k$ satisfies $a_k < b_k$.

For example, $(4, 1, 2, 5) <_{\text{lex}} (4, 1, 3, 0)$ and $(1, 1, 0, 1) <_{\text{lex}} (2, 0, 0, 0)$. The relation $<_{\text{lex}}$ is usually pronounced “is lexicographically smaller than”; the word “lexico-
graphic” comes from the idea that if numbers were letters, then a “word” $a_1 a_2 \cdots a_n$
would appear earlier in a dictionary than $b_1 b_2 \cdots b_n$ if and only if $(a_1, a_2, \ldots, a_n) <_{\text{lex}}$
$(b_1, b_2, \ldots, b_n)$.

Exercise 5. Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ and $\tau \in S_n$. Prove the following:

(a) If

$$(\sigma (1), \sigma (2), \ldots, \sigma (n)) <_{\text{lex}} (\tau (1), \tau (2), \ldots, \tau (n)),$$

then

$$(\ell_1 (\sigma), \ell_2 (\sigma), \ldots, \ell_n (\sigma)) <_{\text{lex}} (\ell_1 (\tau), \ell_2 (\tau), \ldots, \ell_n (\tau)).$$

(b) If $(\ell_1 (\sigma), \ell_2 (\sigma), \ldots, \ell_n (\sigma)) = (\ell_1 (\tau), \ell_2 (\tau), \ldots, \ell_n (\tau))$, then $\sigma = \tau$.

The solution to Exercise 5 given below is one of those cases where a simple
argument becomes insufferably long and dreary as I try to capture it in writing.
Apologies for what you are about to see. The proof relies on the following lemma:
Lemma 0.16. Let \( n \in \mathbb{N} \). Let \( \sigma \in S_n \) and \( i \in [n] \). Then:

(a) We have \( \ell_i(\sigma) = |\{i \mid i+1 < \sigma(i) \} \}. \)

(b) We have \( \ell_i(\sigma) = |\{i \mid i+1 < \sigma(i) \} \}. \)

Proof of Lemma 0.16

(a) We know that \( \ell_i(\sigma) \) is the number of \( j \in \{i+1, i+2, \ldots, n\} \)

such that \( \sigma(i) > \sigma(j) \) (by the definition of \( \ell_i(\sigma) \)). Hence,

\[
\ell_i(\sigma) = \left( \text{the number of } j \in \{i+1, i+2, \ldots, n\} \text{ such that } \sigma(i) > \sigma(j) \right)
\]

\[
\ell_i(\sigma) = |\{j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j) \}| .
\] (17)

Define a set \( A \) by

\[
A = \{ j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j) \} .
\] (18)

Thus,

\[
|A| = |\{ j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j) \}| = \ell_i(\sigma)
\] (by (17)).

Let \( B \) be the set \([\sigma(i) - 1] \setminus \sigma([i])\).

The map \( \sigma \) is a permutation of \([n] \) (since \( \sigma \in S_n \)), and thus is invertible, and therefore is injective.

For each \( k \in A \), we have \( \sigma(k) \in B \). Hence, we can define a map \( \alpha : A \rightarrow B \) by

\[
(\alpha(k) = \sigma(k) \quad \text{for each } k \in A).
\]

Consider this \( \alpha \).

On the other hand, for each \( k \in B \), we have \( \sigma^{-1}(k) \in A \). Hence, we can define a map \( \beta : B \rightarrow A \) by

\[
(\beta(k) = \sigma^{-1}(k) \quad \text{for each } k \in B).
\]

Proof. Let \( k \in A \). Thus, \( k \in A = \{ j \in \{i+1, i+2, \ldots, n\} \mid \sigma(i) > \sigma(j) \} \). In other words, \( k \) is an element of \( \{i+1, i+2, \ldots, n\} \) and satisfies \( \sigma(i) > \sigma(k) \).

From \( k \in \{i+1, i+2, \ldots, n\} \subseteq [n] \), we conclude that \( \sigma(k) \) is well-defined. Also, \( \sigma(k) < \sigma(i) \) (since \( \sigma(i) > \sigma(k) \)), so that \( \sigma(k) \leq \sigma(i) - 1 \) (since \( \sigma(k) \) and \( \sigma(i) \) are integers). Thus, \( \sigma(k) \in [\sigma(i) - 1] \).

Next, let us prove that \( \sigma(k) \notin \sigma([i]) \).

Indeed, assume the contrary (for the sake of contradiction). Hence, \( \sigma(k) \in \sigma([i]) \). In other words, there exists some \( j \in [i] \) such that \( \sigma(k) = \sigma(j) \). Consider this \( j \). From \( \sigma(k) = \sigma(j) \), we obtain \( k = j \) (since the map \( \sigma \) is injective). Hence, \( k = j \in [i] \). But \( k \in \{i+1, i+2, \ldots, n\} = [n] \setminus [i] \), so that \( k \notin [i] \). This contradicts \( k \in [i] \). This contradiction shows that our assumption was false. Hence, \( \sigma(k) \notin \sigma([i]) \) is proven.

Combining \( \sigma(k) \in [\sigma(i) - 1] \) with \( \sigma(k) \notin \sigma([i]) \), we obtain \( \sigma(k) \in [\sigma(i) - 1] \setminus \sigma([i]) = B \).

Qed.
Consider this $\beta$.

The maps $\alpha$ and $\beta$ are mutually inverse (since $\alpha$ is a restriction of $\sigma$, whereas $\beta$ is a restriction of $\sigma^{-1}$), and therefore are bijections. Hence, there is a bijection from $A$ to $B$ (namely, $\alpha$). Thus, $|A| = |B|$.

But (19) yields

$$\ell_i (\sigma) = |A| = |B| = |[\sigma (i) - 1] \setminus \sigma ([i])|$$

(since $B = [\sigma (i) - 1] \setminus \sigma ([i])$). This proves Lemma 0.16 (a).

(b) If we had $\sigma (i) \in [\sigma (i) - 1]$, then we would have $\sigma (i) \leq \sigma (i) - 1 < \sigma (i)$, which would be absurd. Hence, we have $\sigma (i) \notin [\sigma (i) - 1]$.

But $[i] = \{i\} \cup [i - 1]$. Hence,

$$\sigma \left( \begin{array}{c} [i] \\ = \{i\} \cup [i - 1] \end{array} \right) = \sigma \left( \begin{array}{c} \{i\} \cup [i - 1] \\ = \sigma \{i\} \end{array} \right) = \sigma ([i]) \cup [\sigma (i) - 1] \cup [\sigma (i - 1)].$$

Thus,

$$[\sigma (i) - 1] \setminus \sigma ([i]) = [\sigma (i) - 1] \setminus ([\{\sigma (i)\} \cup [\sigma (i - 1)])$$

$$= [\sigma (i) - 1] \setminus ([\{\sigma (i)\} \setminus [\sigma (i)])$$

$$= [\sigma (i) - 1] \setminus [\sigma (i)] = [\sigma (i) - 1] \setminus [\sigma (i - 1)].$$

Now, Lemma 0.16 (a) yields

$$\ell_i (\sigma) = \left| \frac{[\sigma (i) - 1] \setminus \sigma ([i])}{= \sigma (i) - 1, \sigma (i)} \right| = |[\sigma (i) - 1] \setminus [\sigma (i - 1)]| \cdot$$

This proves Lemma 0.16 (b). \qed

**Solution to Exercise 5 (sketched).** (a) Assume that

$$(\sigma (1), \sigma (2), \ldots, \sigma (n)) \preceq_{\text{lex}} (\tau (1), \tau (2), \ldots, \tau (n)).$$

Hence, $k = \sigma \left( \sigma^{-1} (k) \right) \in \sigma ([i])$, which contradicts $k \notin \sigma ([i])$. This contradiction shows that our assumption was false. Thus, $\sigma^{-1} (k) \in \{i + 1, i + 2, \ldots, n\}$ is proven.

Now, we know that $\sigma^{-1} (k) \in \{i + 1, i + 2, \ldots, n\}$ and $\sigma (i) > \sigma (\sigma^{-1} (k))$. In other words, $\sigma^{-1} (k)$ is a $j \in \{i + 1, i + 2, \ldots, n\}$ satisfying $\sigma (i) > \sigma (j)$. In other words,

$$\sigma^{-1} (k) \in \{j \in \{i + 1, i + 2, \ldots, n\} | \sigma (i) > \sigma (j)\}.$$

In view of (18), this rewrites as $\sigma^{-1} (k) \in A$. Qed.
According to Definition \ref{def:smallest}, this means the following: There exists some \( k \in [n] \) such that \( \sigma(k) \neq \tau(k) \), and the \textbf{smallest} such \( k \) satisfies \( \sigma(k) < \tau(k) \).

Let \( i \) be the smallest such \( k \). Thus, \( \sigma(i) < \tau(i) \), but
\[
each k \in [i - 1] \text{ satisfies } \sigma(k) = \tau(k) \tag{20}
\] (since \( i \) is the \textbf{smallest} \( k \in [n] \) such that \( \sigma(k) \neq \tau(k) \)).

Thus,
\[
each k \in [i] \text{ satisfies } \sigma([k - 1]) = \tau([k - 1]) \tag{21}
\]

Hence,
\[
each k \in [i - 1] \text{ satisfies } \ell_k(\sigma) = \ell_k(\tau) \tag{22}
\]

Furthermore,
\[
\ell_i(\sigma) < \ell_i(\tau) \tag{23}
\]

Thus, \( \ell_i(\sigma) \neq \ell_i(\tau) \). In other words, \( i \) is a \( k \in [n] \) such that \( \ell_k(\sigma) \neq \ell_k(\tau) \).

Moreover, \( \text{(22)} \) shows that \( i \) is the \textbf{smallest} such \( k \). Thus, the smallest \( k \in [n] \) such that \( \ell_k(\sigma) \neq \ell_k(\tau) \) satisfies \( \ell_k(\sigma) < \ell_k(\tau) \) (because this \( k \) is \( i \), and \( i \) satisfies \( \text{(23)} \)).

\textbf{Proof of (21)}: Let \( k \in [i] \). Thus, \( k \leq i \).

Let \( j \in [k - 1] \). Thus, \( j \leq k - 1 \leq i - 1 \), so that \( j \in [i - 1] \). Hence, \( (20) \) (applied to \( j \) instead of \( k \)) shows that \( \sigma(j) = \tau(j) \).

Now, forget that we fixed \( j \). We thus have shown that \( \sigma(j) = \tau(j) \) for each \( j \in [k - 1] \). In other words,
\[
(\sigma(1), \sigma(2), \ldots, \sigma(k - 1)) = (\tau(1), \tau(2), \ldots, \tau(k - 1)).
\]

Thus,
\[
\{\sigma(1), \sigma(2), \ldots, \sigma(k - 1)\} = \{\tau(1), \tau(2), \ldots, \tau(k - 1)\}.
\]

Now,
\[
\sigma\left(\begin{array}{c}
 k - 1 \\
 \stackrel{=} {= (1, 2, \ldots, k - 1)}
\end{array}\right) = \sigma\left(\begin{array}{c}
 1, 2, \ldots, k - 1 \\
 \stackrel{=} {= (1, 2, \ldots, k - 1)}
\end{array}\right) = \{\tau(1), \tau(2), \ldots, \tau(k - 1)\} = \tau\left(\begin{array}{c}
 1, 2, \ldots, k - 1 \\
 \stackrel{=} {= (k - 1)}
\end{array}\right) = \tau([k - 1]).
\]

This proves \( \text{(21)} \).

\textbf{Proof of (22)}: Let \( k \in [i - 1] \). Then, \( \text{Lemma \ref{lem:ordering} (b)} \) (applied to \( k \) instead of \( i \)) yields
\[
\ell_k(\sigma) = |\sigma(k) - 1 \setminus \sigma([k - 1])|. \tag{24}
\]

\( \text{The same argument (applied to } \tau \text{ instead of } \sigma \text{) yields}
\[
\ell_k(\tau) = |\tau(k) - 1 \setminus \tau([k - 1])|. \tag{25}
\]

But \( k \in [i - 1] \subseteq [i] \). Hence, \( \text{(21)} \) yields \( \sigma([k - 1]) = \tau([k - 1]) \). Also, \( \text{(20)} \) yields \( \sigma(k) = \tau(k) \).

Hence,
\[
\ell_k(\sigma) = \left|\left|\sigma(k) - 1 \setminus \sigma([k - 1])\right| = |\tau(k) - 1 \setminus \tau([k - 1])| = \ell_k(\tau). \tag{26}
\]

This proves \( \text{(22)} \).

\textbf{Proof of (23)}: We have \( i \in [i] \). Hence, \( \text{(21)} \) (applied to \( k = i \)) yields \( \sigma([i - 1]) = \tau([i - 1]) \).
Thus, we have shown that there exists some $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$, and the **smallest** such $k$ satisfies $\ell_k(\sigma) < \ell_k(\tau)$. But this means precisely that

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) <_{\text{lex}} (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau))$$

(according to Definition 0.15). Hence, Exercise $[n]$ (a) is solved.

(b) Assume that

$$(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) = (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau)).$$

We must prove that $\sigma = \tau$.

Indeed, assume the contrary. Thus, $\sigma \neq \tau$. Hence, there exists some $k \in [n]$ satisfying $\sigma(k) \neq \tau(k)$. Therefore, there exists the **smallest** such $k$. This smallest $k$ must satisfy either $\sigma(k) < \tau(k)$ or $\sigma(k) > \tau(k)$ (because it satisfies $\sigma(k) \neq \tau(k)$).

We can WLOG assume that it satisfies $\sigma(k) < \tau(k)$ (because otherwise, we can simply switch the roles of $\sigma$ and $\tau$). Assume this. Thus, $(\sigma(1), \sigma(2), \ldots, \sigma(n)) <_{\text{lex}} (\tau(1), \tau(2), \ldots, \tau(n))$ (because of Definition 0.15). Hence, Exercise $[n]$ (a) shows that $(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) <_{\text{lex}} (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau))$. In other words, there

Also, $\sigma(i) < \tau(i)$, so that $\sigma(i) - 1 < \tau(i) - 1$ and therefore $[\sigma(i) - 1] \subseteq [\tau(i) - 1]$.

From $\sigma(i) < \tau(i)$, we also obtain $\sigma(i) \leq \tau(i) - 1$ (since $\sigma(i)$ and $\tau(i)$ are integers), and thus $\sigma(i) \in [\tau(i) - 1]$. Also, $\sigma(i) \notin [\sigma(i) - 1]$. [Proof: Assume the contrary. Thus, $\sigma(i) \in [\sigma(i) - 1]$. In other words, $\sigma(i) = \sigma(j)$ for some $j \in [\sigma(i) - 1]$. Consider this $j$. From $\sigma(i) = \sigma(j)$, we obtain $i = j$ (since $\sigma$ is injective), so that $i = j \in [\sigma(i) - 1]$ and thus $i < \sigma(i) - 1$. But this is absurd. Hence, we found a contradiction, so that $\sigma(i) \notin [\sigma(i) - 1]$ is proven.]

If we had $\sigma(i) \in [\sigma(i) - 1]$, then we would have $\sigma(i) \leq \sigma(i) - 1 < \sigma(i)$, which is absurd.

Hence, we have $\sigma(i) \notin [\sigma(i) - 1]$. Thus, also $\sigma(i) \notin [\sigma(i) - 1] \setminus [\sigma(i) - 1]$.

Combining $\sigma(i) \in [\tau(i) - 1]$ with $\sigma(i) \notin [\sigma(i) - 1]$, we obtain $\sigma(i) \in [\tau(i) - 1] \setminus [\sigma(i) - 1]$.

Now,

\[
[\sigma(i) - 1] \setminus [\sigma(i) - 1] \subseteq [\tau(i) - 1] \setminus [\sigma(i) - 1].
\]

Moreover, the set $[\tau(i) - 1] \setminus [\sigma(i) - 1]$ contains $\sigma(i)$ (since $\sigma(i) \in [\tau(i) - 1] \setminus [\sigma(i) - 1]$), but the set $[\tau(i) - 1] \setminus [\sigma(i) - 1]$ does not (since $\sigma(i) \notin [\tau(i) - 1] \setminus [\sigma(i) - 1]$). Thus, these two sets are distinct. In other words, $[\tau(i) - 1] \setminus [\sigma(i) - 1] \neq [\tau(i) - 1] \setminus [\sigma(i) - 1]$. Combining this with $[24]$, we conclude that $[\tau(i) - 1] \setminus [\sigma(i) - 1]$ is a **proper** subset of $[\tau(i) - 1] \setminus [\sigma(i) - 1]$.

Thus,

\[
||\sigma(i) - 1| \setminus \tau(i) - 1| < ||\tau(i) - 1| \setminus \sigma(i) - 1||
\]

(since a proper subset of any finite set must always have smaller size than the latter).

But Lemma 0.16 (b) yields $\ell_i(\sigma) = ||\sigma(i) - 1| \setminus \sigma(i) - 1||$. The same argument (applied to $\tau$ instead of $\sigma$) yields $\ell_i(\tau) = ||\tau(i) - 1| \setminus \tau(i) - 1||$. Hence,

\[
\ell_i(\sigma) = ||\sigma(i) - 1| \setminus \sigma(i) - 1|| < ||\tau(i) - 1| \setminus \tau(i) - 1|| = \ell_i(\tau).
\]

This proves $[23]$. 


exists some $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$, and the smallest such $k$ satisfies $\ell_k(\sigma) < \ell_k(\tau)$ (according to Definition 0.15).

In particular, there exists some $k \in [n]$ such that $\ell_k(\sigma) \neq \ell_k(\tau)$. In other words, $(\ell_1(\sigma), \ell_2(\sigma), \ldots, \ell_n(\sigma)) \neq (\ell_1(\tau), \ell_2(\tau), \ldots, \ell_n(\tau))$. But this contradicts (25). This contradiction shows that our assumption was false. Hence, $\sigma = \tau$ is proven. This solves Exercise 5(b). \qed

0.6. Comparing subsets of $[n]$

If $I$ and $J$ are two finite sets of integers, then we write $I \leq\# J$ if and only if the following two properties hold:

- We have $|I| \geq |J|$.
- For every $r \in \{1, 2, \ldots, |J|\}$, the $r$-th smallest element of $I$ is $\leq$ to the $r$-th smallest element of $J$.

For example, $\{2, 4\} \leq\# \{2, 5\}$ and $\{1, 3\} \leq\# \{2, 4\}$ and $\{1, 3, 5\} \leq\# \{2, 4\}$. (But not $\{1, 3\} \leq\# \{2, 4, 5\}$.)

The relation $\leq\#$ is called the Gale order on the powerset of $[n]$.

**Exercise 6.** Let $n \in \mathbb{N}$. Let $I$ and $J$ be two subsets of $[n]$.

(a) For every subset $S$ of $[n]$ and every $\ell \in [n]$, let $\alpha_S(\ell)$ denote the number of all elements of $S$ that are $\leq \ell$. Prove that $I \leq\# J$ holds if and only if every $\ell \in [n]$ satisfies $\alpha_I(\ell) \geq \alpha_J(\ell)$.

(b) Prove that $I \leq\# J$ if and only if $[n] \setminus J \leq\# [n] \setminus I$.

The following solution is mostly copypasted from [GriRei18 Proof of Proposition 12.75.2], where the exercise serves as a lemma for a combinatorial proof of an identity between Schur polynomials.

**Solution to Exercise 6** (a) We must prove the equivalence

\[ (I \leq\# J) \iff (\text{every } \ell \in [n] \text{ satisfies } \alpha_I(\ell) \geq \alpha_J(\ell)) \quad (26) \]

$\implies$: Assume that $I \leq\# J$. In other words, the following two properties hold:

Property $\alpha$: We have $|I| \geq |J|$.

Property $\beta$: For every $r \in \{1, 2, \ldots, |J|\}$, the $r$-th smallest element of $I$ is $\leq$ to the $r$-th smallest element of $J$.

Now, let $\ell \in [n]$. Then, we need to show that $\alpha_I(\ell) \geq \alpha_J(\ell)$. Since this is obvious if $\alpha_J(\ell) = 0$ (because $\alpha_I(\ell) \geq 0$), we can WLOG assume that $\alpha_J(\ell) \neq 0$. Assume this. Thus, $\alpha_J(\ell) \geq 1$. Also, $\alpha_J(\ell) = \sum_{s \in J \setminus \ell \subseteq J} s \leq \ell \leq |J| \leq |I|$ (since $|I| \geq |J|$).
Hence, both the \(a_J(\ell)\)-th smallest element of \(J\) and the \(a_J(\ell)\)-th smallest element of \(I\) are well-defined.

Since \(a_J(\ell) = |\{s \in J \mid s \leq \ell\}|\), we know that the elements of \(J\) which are \(\leq \ell\) are precisely the \(a_J(\ell)\) smallest elements of \(J\). Thus,

\[(\text{the } a_J(\ell)-\text{th smallest element of } J) = (\text{the largest element of } J \text{ which is } \leq \ell).\]

But by Property \(\beta\) (applied to \(r = a_J(\ell)\)), we have

\[(\text{the } a_J(\ell)-\text{th smallest element of } I) \leq (\text{the } a_J(\ell)-\text{th smallest element of } J) = (\text{the largest element of } J \text{ which is } \leq \ell) \leq \ell.

Hence, there are at least \(a_J(\ell)\) elements of \(I\) which are \(\leq \ell\) (namely, the \(a_J(\ell)\) smallest ones). In other words, \(|\{s \in I \mid s \leq \ell\}| \geq a_J(\ell)\). Now, the definition of \(a_I(\ell)\) yields \(a_I(\ell) = |\{s \in I \mid s \leq \ell\}| \geq a_J(\ell)\). We thus have proven the \(\implies\) direction of (26).

\(\iff\): Assume that every \(\ell \in [n]\) satisfies \(a_I(\ell) \geq a_J(\ell)\). We need to prove that \(I \leq \# J\). In other words, we need to prove that the following two properties hold:

\[\text{Property } \alpha:\ \text{We have } |I| \geq |J|.

\[\text{Property } \beta:\ \text{For every } r \in \{1, 2, \ldots, |J|\}, \text{ the } r\text{-th smallest element of } I \text{ is } \leq \text{ to the } r\text{-th smallest element of } J.

First of all, \(\{s \in I \mid s \leq n\} = I\) (since every \(s \in I\) satisfies \(s \leq n\)), and the definition of \(a_I(n)\) yields \(a_I(n) = \left\lfloor \frac{|s \in I \mid s \leq n|}{|I|} \right\rfloor = |I|\). Similarly, \(a_J(n) = |J|\).

Applying \(a_I(\ell) \geq a_J(\ell)\) to \(\ell = n\), we obtain \(a_I(n) \geq a_J(n)\), so that \(|I| = a_I(n) \geq a_J(n) = |J|\), and thus Property \(\alpha\) is proven.

Now, let \(r \in \{1, 2, \ldots, |J|\}\). The \(r\)-th smallest element of \(I\) and the \(r\)-th smallest element of \(J\) are then well-defined (because of \(r \leq |J| \leq |I|\)). Let \(\ell\) be the \(r\)-th smallest element of \(J\). Then, \(\{s \in J \mid s \leq \ell\}\) is the set consisting of the \(r\) smallest elements of \(J\), so that \(|\{s \in J \mid s \leq \ell\}| = r\). Now, the definition of \(a_J(\ell)\) yields \(a_J(\ell) = |\{s \in J \mid s \leq \ell\}| = r\).

But the definition of \(a_I(\ell)\) yields \(a_I(\ell) = |\{s \in I \mid s \leq \ell\}|\), so that

\(|\{s \in I \mid s \leq \ell\}| = a_I(\ell) \geq a_J(\ell) = r\).

In other words, there exist at least \(r\) elements of \(I\) which are \(\leq \ell\). Hence, the \(r\)-th smallest element of \(I\) must be \(\leq \ell\). Since \(\ell\) is the \(r\)-th smallest element of \(J\), this rewrites as follows: The \(r\)-th smallest element of \(I\) is \(\leq\) to the \(r\)-th smallest element of \(J\). Thus, Property \(\beta\) holds. Now we know that both Properties \(\alpha\) and \(\beta\) hold. Hence, \(I \leq \# J\) holds (which, as we know, is equivalent to the conjunction of said properties). This proves the \(\iff\) direction of (26). Thus, (26) is proven. In other words, Exercise [6](a) is solved.
(b) For every $\ell \in [n]$ and $S \subseteq [n]$, let $\alpha_S(\ell)$ denote the number $|\{s \in S \mid s \leq \ell\}|$. Thus, every $\ell \in [n]$ satisfies
\[
\alpha_I(\ell) + \alpha_{[n] \setminus I}(\ell) = |\{s \in I \mid s \leq \ell\}| + |\{s \in [n] \setminus I \mid s \leq \ell\}|
\]
\[
= \left| \left\{ s \in I \cup ([n] \setminus I) \mid s \leq \ell \right\} \right| = |\{s \in [n] \mid s \leq \ell\}| = |\{1, 2, \ldots, \ell\}| = \ell,
\]
so that $\alpha_{[n] \setminus I}(\ell) = \ell - \alpha_I(\ell)$. Similarly, every $\ell \in [n]$ satisfies $\alpha_{[n] \setminus J}(\ell) = \ell - \alpha_J(\ell)$.
Applying (26) to $[n] \setminus J$ and $[n] \setminus I$ in lieu of $I$ and $J$, we obtain the equivalence
\[
([n] \setminus J \leq \# [n] \setminus I) \iff (\text{every } \ell \in [n] \text{ satisfies } \alpha_{[n] \setminus J}(\ell) \geq \alpha_{[n] \setminus I}(\ell))
\]
Hence, we have the following chain of equivalences:
\[
([n] \setminus I \leq \# [n] \setminus I)
\]
\[
\iff (\text{every } \ell \in [n] \text{ satisfies } \alpha_{[n] \setminus J}(\ell) \geq \alpha_{[n] \setminus I}(\ell) = \ell - \alpha_I(\ell))
\]
\[
\iff (\text{every } \ell \in [n] \text{ satisfies } \ell - \alpha_I(\ell) \geq \ell - \alpha_J(\ell))
\]
\[
\iff (\text{every } \ell \in [n] \text{ satisfies } \alpha_I(\ell) \geq \alpha_J(\ell))
\]
\[
\iff ([I \leq \# J]) \quad \text{(by (26))}
\]
This solves Exercise 6(b). \qed

Remark 0.17. Recall that we have defined a Dyck word as a list $w$ of $2n$ numbers, exactly $n$ of which are 0’s while the other $n$ are 1’s, and having the property that for each $n \in [2n]$, the number of 0’s among the first $k$ entries of $w$ is $\leq$ to the number of 1’s among the first $k$ entries of $w$.

It is not hard to see the connection between the relation $\leq \#$ and Dyck words: Let $w = (w_1, w_2, \ldots, w_{2n}) \in \{0, 1\}^{2n}$ be a list of $2n$ numbers, exactly $n$ of which are 0’s while the other $n$ are 1’s. Then, $w$ is a Dyck word if and only if
\[
\{i \in [2n] \mid w_i = 1\} \leq \# \{i \in [2n] \mid w_i = 0\}
\]
(in other words, for every $r \in [n]$, the $r$-th appearance of 1 in $w$ precedes the $r$-th appearance of 0 in $w$).

0.7. A rigorous approach to the existence of a cycle decomposition

The purpose of the following exercise is to give a rigorous proof of the fact that any permutation can be decomposed into disjoint cycles.
Exercise 7. Let \( X \) be a finite set. Let \( \sigma \) be a permutation of \( X \).

Define a binary relation \( \sim \) on the set \( X \) as follows: For two elements \( x \in X \) and \( y \in X \), we set \( x \sim y \) if and only if there exists some \( k \in \mathbb{N} \) such that \( y = \sigma^k(x) \).

(a) Prove that \( \sim \) is an equivalence relation.

For any \( x \in X \), we let \( [x]_\sim \) denote the \( \sim \)-equivalence class of \( x \).

(b) For any \( x \in X \), prove that \( [x]_\sim = \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)\} \), where \( k = |[x]_\sim| \).

(c) For any \( \sim \)-equivalence class \( E \), let us define \( c_E \) to be the map

\[
X \to X, \quad x \mapsto \begin{cases} 
\sigma(x), & \text{if } x \in E, \\
x, & \text{if } x \notin E.
\end{cases}
\]

Prove that \( c_E \) is a permutation of \( X \).

(d) Prove that if \( E = [x]_\sim \) for some \( x \in X \), then \( c_E \) can be written as \( \text{cyc}_{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)} \), where \( k = |[x]_\sim| \). (Don't forget to show that \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \) are distinct, so that \( \text{cyc}_{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)} \) is well-defined.)

(e) Let \( E_1, E_2, \ldots, E_m \) be all \( \sim \)-equivalence classes (listed without repetitions – that is, \( E_i \neq E_j \) whenever \( i \neq j \)). Prove that

\[
\sigma = c_{E_1} \circ c_{E_2} \circ \cdots \circ c_{E_m}.
\]

Exercise 7 is mostly an exercise in understanding the definitions and writing up proofs. The first two parts of it are similar to Exercise 6 on homework set #3; thus, our solution below is partly copypasted from the latter (with the necessary changes made).

Our solution relies on a few lemmas:

Lemma 0.18. Let \( X \) be a set. Let \( f : X \to X \) be any map. Let \( x \in X \). Let \( i \) and \( j \) be two nonnegative integers satisfying \( i < j \) and \( f^i(x) = f^j(x) \). Then,

\[
\left\{ f^h(x) \mid h \in \mathbb{N} \right\} = \left\{ f^0(x), f^1(x), \ldots, f^{j-1}(x) \right\}.
\]

Proof of Lemma 0.18. We have

\[
\left\{ f^0(x), f^1(x), \ldots, f^{j-1}(x) \right\} = \left\{ f^h(x) \mid h \in \{0, 1, \ldots, j-1\} \right\} \subseteq \left\{ f^h(x) \mid h \in \mathbb{N} \right\}. \tag{27}
\]

On the other hand, we have \( i \in \{0, 1, \ldots, j-1\} \) (since \( i \) is a nonnegative integer satisfying \( i < j \)), and thus \( f^i(x) \in \{ f^0(x), f^1(x), \ldots, f^{j-1}(x) \} \). Hence, \( \{ f^i(x) \} \subseteq \)
{f^0(x), f^1(x), \ldots, f^{i-1}(x)}. Therefore,

\[
\{f^0(x), f^1(x), \ldots, f^{i-1}(x)\} = \left\{f^0(x), f^1(x), \ldots, f^{i-1}(x)\right\} \cup \left\{f^i(x) \atop = f^j(x)\right\}
\]

\[
= \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\} \cup \{f^i(x)\}
\]

\[
= \{f^0(x), f^1(x), \ldots, f^i(x)\}.
\]

Now,

\[
f^h(x) \in \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\} \quad \text{for each } h \in \mathbb{N}.
\]

Proof of (29): We shall prove (29) by induction over \(h\):

Induction base: We have \(i < j\), hence \(j > i \geq 0\) and thus \(j \geq 1\) (since \(j\) is an integer). Hence, \(0 \in \{0, 1, \ldots, j-1\}\), so that \(f^0(x) \in \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\}\). In other words, (29) holds for \(h = 0\). This completes the induction base.

Induction step: Let \(g \in \mathbb{N}\). Assume that (29) holds for \(h = g\). We must now show that (29) holds for \(h = g + 1\) as well.

We have assumed that (29) holds for \(h = g\). In other words, \(f^g(x) \in \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\}\). In other words, there exists some \(k \in \{0, 1, \ldots, j-1\}\) such that \(f^g(x) = f^k(x)\). Consider this \(k\).

We have \(k \in \{0, 1, \ldots, j-1\}\), so that \(k + 1 \in \{1, 2, \ldots, j\} \subseteq \{0, 1, \ldots, j\}\) and therefore

\[
f^{k+1}(x) \in \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\}
\]

(by (28)). But

\[
f^{g+1}(x) = f \left( f^g(x) \atop = f^k(x)\right) = f \left( f^g(x) \right) = f^{k+1}(x) \in \{f^0(x), f^1(x), \ldots, f^{i-1}(x)\}
\]

(as we have just proven). In other words, (29) holds for \(h = g + 1\) as well. This completes the induction step. Thus, (29) is proven.

From (29), we immediately obtain

\[
\left\{f^h(x) \mid h \in \mathbb{N}\right\} \subseteq \left\{f^0(x), f^1(x), \ldots, f^{i-1}(x)\right\}.
\]

Combining this with (27), we obtain

\[
\left\{f^h(x) \mid h \in \mathbb{N}\right\} = \left\{f^0(x), f^1(x), \ldots, f^{i-1}(x)\right\}.
\]

This proves Lemma 0.18.

\[\square\]

Lemma 0.19. Let \(X\) be a finite set. Let \(\sigma\) be a permutation of \(X\). Let \(x \in X\).

(a) There exists a \(j \in \mathbb{N}\) such that \(\sigma^j(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{j-1}(x)\}\).

Let \(p\) be the smallest such \(j\).

(b) The integer \(p\) is positive and satisfies \(\sigma^p(x) = x\).

(c) The elements \(\sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x)\) are pairwise distinct.

(d) We have \(\{\sigma^h(x) \mid h \in \mathbb{N}\} = \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x)\}\).
Proof of Lemma 0.19. The map \( \sigma \) is a permutation of \( X \). In other words, \( \sigma \) is a bijection \( X \to X \). Hence, \( \sigma \) is injective.

(a) Define an \( n \in \mathbb{N} \) by \( n = |X| \). (This is well-defined, since \( X \) is a finite set.) The \( n + 1 \) elements \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^n(x) \) cannot all be distinct, because they all belong to the \( n \)-element set \( X \). Hence, at least two of these \( n + 1 \) elements are equal. In other words, there exist two elements \( u, v \) of \( \{0, 1, \ldots, n\} \) such that \( u < v \) and \( \sigma^u(x) = \sigma^v(x) \). Consider these \( u \) and \( v \).

We have \( u \in \{0, 1, \ldots, n\} \subseteq \mathbb{N} \). Thus, \( u \in \{0, 1, \ldots, v - 1\} \) (since \( u < v \)). Hence, 
\[
\sigma^u(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{v-1}(x)\}. 
\]
In view of \( \sigma^u(x) = \sigma^v(x) \), this rewrites as 
\[
\sigma^v(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{v-1}(x)\}. 
\]
Thus, there exists a \( j \in \mathbb{N} \) such that 
\[
\sigma^j(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{i-1}(x)\} \tag{namely, \( j = v \)). 
\]
This proves Lemma 0.19 (a).

Now, let us study the \( p \) in Lemma 0.19. We have defined \( p \) as the smallest \( j \in \mathbb{N} \) such that 
\[
\sigma^j(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x)\}. 
\]
In other words, there exists some \( i \in \{0, 1, \ldots, p - 1\} \) such that 
\[
\sigma^p(x) = \sigma^i(x). 
\]
Consider this \( i \).

Next, we claim that \( i = 0 \). We shall prove this by contradiction. Indeed, assume the contrary. Thus, \( i \neq 0 \), so that \( i > 0 \) (since \( i \in \mathbb{N} \)). Hence, 
\[
\sigma^i(x) = \sigma^{i-1}(x). 
\]
But the integer \( p \) is also positive; hence, \( p - 1 \in \mathbb{N} \) and \( \sigma^p(x) = \sigma^{p-1}(x) \). Hence, 
\[
\sigma^{p-1}(x) = \sigma^{p}(x) = \sigma^{i-1}(x). 
\]
Since \( \sigma \) is injective, we thus conclude that 
\[
\sigma^{-1}(x) = \sigma^{-1}(x). 
\]
But \( i - 1 \in \mathbb{N} \) (since \( i > 0 \)). From \( i \in \{0, 1, \ldots, p - 1\} \), we obtain \( i - 1 \in \{-1, 0, \ldots, (p - 1) - 1\} \). Combined with \( i - 1 \in \mathbb{N} \), this yields 
\[
i - 1 \in \{-1, 0, \ldots, (p - 1) - 1\} \cap \mathbb{N} = \{0, 1, \ldots, (p - 1) - 1\}. 
\]
Hence, 
\[
\sigma^{i-1}(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{(p-1)-1}(x)\}. 
\]
Hence, 
\[
\sigma^{p-1}(x) = \sigma^{i-1}(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{(p-1)-1}(x)\}. 
\]
Thus, \( p - 1 \) is a \( j \in \mathbb{N} \) such that 
\[
\sigma^j(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{j-1}(x)\} \tag{because \( p - 1 \in \mathbb{N} \)). 
\]
But we defined \( p \) to be the smallest such \( j \). Hence, \( p \leq p - 1 \). This contradicts \( p > p - 1 \). This contradiction shows that our assumption was false; hence, we have shown that \( i = 0 \). Therefore, 
\[
\sigma^i(x) = \underbrace{\sigma^0(x)}_{= \text{id}} = \text{id}(x) = x. 
\]

(c) Assume the contrary. Thus, two of the elements \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x) \) are equal. In other words, there exist two elements \( u, v \) of \( \{0, 1, \ldots, p - 1\} \) such that \( u < v \) and \( \sigma^u(x) = \sigma^v(x) \). Consider these \( u \) and \( v \). Notice that \( v \leq p - 1 \) (since \( v \in \{0, 1, \ldots, p - 1\} \)).

From \( u \in \{0, 1, \ldots, p - 1\} \), we obtain \( u \geq 0 \). From \( u < v \), we obtain \( u \leq v - 1 \) (since \( u, v \) are integers), so that \( u \in \{0, 1, \ldots, v - 1\} \) (since \( u \geq 0 \)).
Hence, $\sigma^u(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{u-1}(x)\}$. From $\sigma^u(x) = \sigma^v(x)$, we obtain $\sigma^v(x) = \sigma^u(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{v-1}(x)\}$. Thus, $v$ is a $j \in \mathbb{N}$ such that $\sigma^j(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{j-1}(x)\}$ (because $v \in \mathbb{N}$). But we defined $p$ to be the smallest such $j$. Hence, $p \leq v$. This contradicts $v \leq p - 1 < p$. This contradiction shows that our assumption was false. Thus, Lemma 0.19(c) is proven.

(d) We have $\sigma^p(x) \in \{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x)\}$. In other words, there exists some $i \in \{0, 1, \ldots, p - 1\}$ such that $\sigma^p(x) = \sigma^i(x)$. Consider this $i$. Hence, $i < p$ (since $i \in \{0, 1, \ldots, p - 1\}$) and $\sigma^j(x) = \sigma^p(x)$. Thus, Lemma 0.18 (applied to $f = \sigma$ and $j = p$) yields

$$\left\{\sigma^h(x) \mid h \in \mathbb{N}\right\} = \left\{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x)\right\}.$$ 

This proves Lemma 0.19(d). \qed

Lemma 0.20. Let $X$ be a set. Let $f : X \to X$ be any map. Let $x \in X$. Let $p \in \mathbb{N}$ be such that $f^p(x) = x$. Then, $f^{kp}(x) = x$ for each $k \in \mathbb{N}$.

Proof of Lemma 0.20. Lemma 0.20 is intuitively obvious: All it says is that if applying the map $f$ to $x$ a total of $p$ times brings you back to $x$, then applying the map $f$ to $x$ a total of $kp$ times brings you back to $x$ as well. This intuition can easily be translated into a rigorous argument:

We shall prove Lemma 0.20 by induction over $k$:

Induction base: We have $f^0 = f^0 = \text{id}_X$, so that $f^0(x) = \text{id}_X(x) = x$. Thus, Lemma 0.20 holds for $k = 0$. This completes the induction base.

Induction step: Let $m \in \mathbb{N}$. Assume that Lemma 0.20 holds for $k = m$. We must prove that Lemma 0.20 holds for $k = m + 1$.

Let $x \in X$. Let $p \in \mathbb{N}$ be such that $f^p(x) = x$. Then, $f^{mp}(x) = x$ (since Lemma 0.20 holds for $k = m$). But $f^{mp} \circ f^p = f^{mp+p} = f^{(m+1)p}$. Hence, $(f^{mp} \circ f^p)(x) = f^{(m+1)p}(x)$, and therefore

$$f^{(m+1)p}(x) = (f^{mp} \circ f^p)(x) = f^{mp}\left(f^p(x) = x\right) = f^{mp}(x) = x.$$ 

In other words, Lemma 0.20 holds for $k = m + 1$. This completes the induction step. Thus, Lemma 0.20 is proven. \qed

Lemma 0.21. Let $X$ be a set. Let $m \in \mathbb{N}$. Let $f_1, f_2, \ldots, f_m$ be $m$ maps from $X$ to $X$. Let $x$ and $y$ be two elements of $X$.

Let $i \in [m]$. Assume that $f_i(x) = y$. Assume further that

$$f_j(x) = x \quad \text{for each } j \in [m] \text{ satisfying } j < i. \quad (30)$$

Assume also that

$$f_j(y) = y \quad \text{for each } j \in [m] \text{ satisfying } j > i. \quad (31)$$

Then, $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y$. 

Proof of Lemma 0.21: The idea behind this proof is very simple (if we don’t insist on being rigorous): Imagine the element \( x \) undergoing the maps \( f_1, f_2, \ldots, f_m \) in this order; the result is, of course, \( (f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) \). But let us look closer at the step-by-step procedure. The element is initially \( x \). Then, the maps \( f_1, f_2, \ldots, f_m \) are being applied to it in this order. Up until the map \( f_i \) is applied, the element does not change (because of (30)). Then, the map \( f_i \) is applied, and the element becomes \( y \) (since \( f_i(x) = y \)). From then on, the maps \( f_{i+1}, f_{i+2}, \ldots, f_m \) again leave the element unchanged (due to (31)). Thus, the final result is \( y \). This shows that \( (f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y \).

Let us now rewrite the above argument in rigorous terms.

We have \( i \in [m] \), so that \( 1 \leq i \leq m \). Now, we claim the following:

**Observation 1:** We have \( (f_g \circ f_{g-1} \circ \cdots \circ f_1)(x) = \begin{cases} x, & \text{if } g < i; \\ y, & \text{if } g \geq i \end{cases} \) for each \( g \in \{0, 1, \ldots, m\} \).

**Proof of Observation 1:** We shall prove Observation 1 by induction on \( g \):

**Induction base:** We have \( 0 < i \) (since \( i \in [m] \)). Thus, \( \begin{cases} x, & \text{if } 0 < i; \\ y, & \text{if } 0 \geq i \end{cases} = x \). Comparing this with \( (f_0 \circ f_{0-1} \circ \cdots \circ f_1)(x) = \text{id}(x) = x \),

we obtain \( (f_0 \circ f_{0-1} \circ \cdots \circ f_1)(x) = \begin{cases} x, & \text{if } 0 < i; \\ y, & \text{if } 0 \geq i \end{cases} \). In other words, Observation 1 holds for \( g = 0 \).

This completes the induction base.

**Induction step:** Let \( h \in \{0, 1, \ldots, m\} \) be positive. Assume that Observation 1 holds for \( g = h - 1 \). We must then prove that Observation 1 holds for \( g = h \).

We have

\[
(f_h \left( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} \right) = \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases}) \tag{32}
\]

Proof of (32): We are in one of the following three cases:

**Case 1:** We have \( h < i \).

**Case 2:** We have \( h = i \).

**Case 3:** We have \( h > i \).

Let us first consider Case 1. In this case, we have \( h < i \). Thus, \( \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases} = x \).

Applying (30) to \( j = h \), we find \( f_h(x) = x \) (since \( h < i \)).

Also, \( h - 1 < h < i \). Hence, \( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} = x \). Applying the map \( f_h \) to this equality, we obtain

\[
f_h \left( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} \right) = f_h(x) = x = \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases}.
\]

Hence, (32) is proven in Case 1.

Let us now consider case 2. In this case, we have \( h = i \). Thus, \( h \geq i \). Hence, \( \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases} = y \).
But we assumed that Observation 1 holds for $g = h - 1$. In other words, we have $(f_{h-1} \circ f_{(h-1)-1} \circ \cdots \circ f_1)(x) = \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases}$.

Now,

$$
(f_h \circ f_{h-1} \circ \cdots \circ f_1)(x) = f_h(f_{h-1} \circ f_{(h-1)-1} \circ \cdots \circ f_1)(x) = f_h \left( f_{h-1} \circ f_{(h-1)-1} \circ \cdots \circ f_1 \right)(x) = f_h \left( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} \right) = \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases} \quad \text{(by (32)).}
$$

In other words, Observation 1 holds for $g = h$. This completes the induction step. Thus, Observation 1 is proven.

We can now apply Observation 1 to $g = m$. We thus obtain

$$(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = \begin{cases} x, & \text{if } m < i; \\ y, & \text{if } m \geq i \end{cases} = y$$

(since $m \geq i$ (since $i \leq m$)). This proves Lemma 0.21.

From $h = i$, we obtain $f_h(x) = f_i(x) = y$.

Also, $h - 1 < h = i$. Hence, $f_h(x) = y = x$. Applying the map $f_h$ to this equality, we obtain

$$f_h \left( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} \right) = f_h(x) = y = \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases}.$$

Hence, (32) is proven in Case 2.

Let us first consider Case 3. In this case, we have $h > i$. Thus, $h \geq i$, so that $f_h(x) = y$.

Applying (31) to $j = h$, we find $f_h(y) = y$ (since $h > i$).

Also, $h > i$, so that $h \geq i + 1$ (since $h$ and $i$ are integers). Thus, $h - 1 \geq i$. Hence, $f_h(y) = y$. Applying the map $f_h$ to this equality, we obtain

$$f_h \left( \begin{cases} x, & \text{if } h - 1 < i; \\ y, & \text{if } h - 1 \geq i \end{cases} \right) = f_h(y) = y = \begin{cases} x, & \text{if } h < i; \\ y, & \text{if } h \geq i \end{cases}.$$

Hence, (32) is proven in Case 3.

We have now proven (32) in each of the three Cases 1, 2 and 3 (which are the only cases that can occur). Thus, (32) always holds.
Lemma 0.22. Let $X$ be a set. Let $m \in \mathbb{N}$. Let $g_1, g_2, \ldots, g_m$ be $m$ maps from $X$ to $X$. Let $x$ and $y$ be two elements of $X$.

Let $i \in [m]$. Assume that $g_i(x) = y$. Assume further that
\[
g_j(x) = x \quad \text{for each } j \in [m] \text{ satisfying } j > i.
\] (33)

Assume also that
\[
g_j(y) = y \quad \text{for each } j \in [m] \text{ satisfying } j < i.
\] (34)

Then, \((g_1 \circ g_2 \circ \cdots \circ g_m)(x) = y\).

Proof of Lemma 0.22 Lemma 0.22 follows by applying Lemma 0.21 to $g_m, g_{m-1}, \ldots, g_1$ instead of $f_1, f_2, \ldots, f_m$.

Here is the argument in more detail:

For each $j \in [m]$, we define a map $f_j$ from $X$ to $X$ by $f_j = g_{m+1-j}$.

From $i \in [m]$, we obtain $m + 1 - i \in [m]$. Thus, we can define $i' \in [m]$ by $i' = m + 1 - i$.

Consider this $i'$. From $i' = m + 1 - i$, we obtain $m + 1 - i' = i$. Now, the definition of $f_{i'}$ yields $f_{i'} = g_{m+1-i'} = g_i$ (since $m + 1 - i' = i$). Thus, $f_{i'}(x) = g_i(x) = y$.

Furthermore, $f_j(x) = x$ for each $j \in [m]$ satisfying $j < i'$ \footnote{17}. Also, $f_j(y) = y$ for each $j \in [m]$ satisfying $j > i'$ \footnote{18}. Hence, Lemma 0.21 applied to $i'$ instead of $i$ yields $(f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y$.

But each $j \in [m]$ satisfies
\[
f_{m+1-j} = g_{m+1-(m+1-j)} \quad \text{(by the definition of } f_{m+1-j})
\]
\[
= g_j \quad \text{(since } m + 1 - (m + 1 - j) = j)\]

In other words, we have $f(m, f_{m-1}, \ldots, f_1) = (g_1, g_2, \ldots, g_m)$. Hence, $f_m \circ f_{m-1} \circ \cdots \circ f_1 = g_1 \circ g_2 \circ \cdots \circ g_m$. Hence, $f_m \circ f_{m-1} \circ \cdots \circ f_1(x) = (g_1 \circ g_2 \circ \cdots \circ g_m)(x)$. Therefore,
\[
(g_1 \circ g_2 \circ \cdots \circ g_m)(x) = (f_m \circ f_{m-1} \circ \cdots \circ f_1)(x) = y.
\]

This proves Lemma 0.22 \footnote{\textit{Solution to Exercise 7} (sketched). The map $\sigma$ is a permutation of $X$, thus a bijection $X \to X$. Hence, in particular, $\sigma$ is injective.

Before we properly start solving the exercise, let us make some basic observations:

Observation 1. For every $x \in X$, there exists some positive integer $p$ such that $\sigma^p(x) = x$.}

Proof. Let $j \in [m]$ be such that $j < i'$. Then, $m + 1 - j \in [m]$ (since $j \in [m]$) and $m + 1 - j > m + 1 - i' = i$. Hence, (33) (applied to $m + 1 - j$ instead of $j$) yields $g_{m+1-j}(x) = x$. But the definition of $f_j$ yields $f_j = g_{m+1-j}$. Thus, $f_j(x) = g_{m+1-j}(x) = x$. Qed.

Proof. Let $j \in [m]$ be such that $j > i'$. Then, $m + 1 - j \in [m]$ (since $j \in [m]$) and $m + 1 - j < m + 1 - i' = i$. Hence, (34) (applied to $m + 1 - j$ instead of $j$) yields $g_{m+1-j}(y) = y$. But the definition of $f_j$ yields $f_j = g_{m+1-j}$. Thus, $f_j(y) = g_{m+1-j}(y) = y$. Qed.
[Proof of Observation 1:] Let \( x \in X \). Let \( n = |X| \). The \( n + 1 \) elements \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^n(x) \) cannot all be distinct, because they belong to the \( n \)-element set \( X \). Hence, at least two of these \( n + 1 \) elements are equal. In other words, there exist two elements \( i \) and \( j \) of \( \{0, 1, \ldots, n\} \) such that \( i < j \) and \( \sigma^i(x) = \sigma^j(x) \). Consider these \( i \) and \( j \). From \( i < j \), we conclude that \( j - i \) is a positive integer. Thus, \( \sigma^i = \sigma^j \circ \sigma^{j-i} \).

Also, the map \( \sigma^i \) is injective (since the map \( \sigma \) is injective, but any composition of injective maps is injective). Hence, from

\[
\sigma^i(x) = \sum_{i=0}^{j-i} (\sigma^j \circ \sigma^{j-i})(x) = \sigma^j \left( \sigma^{j-i}(x) \right),
\]

we obtain \( x = \sigma^{j-i}(x) \). In other words, \( \sigma^{j-i}(x) = x \). Hence, there exists some positive integer \( p \) such that \( \sigma^p(x) = x \) (namely, \( p = j - i \)). This proves Observation 1.

Now, we must show that \( \sim \) is an equivalence relation. Indeed, the relation \( \sim \) is reflexive, symmetric, and transitive. In other words, the relation \( \sim \) is an equivalence relation. This solves Exercise 7(a).

19 Proof. Let \( x \in X \). We shall show that \( x \sim x \).

Indeed, \( \sigma^0 = \text{id}_X \), so that \( \sigma^0(x) = \text{id}_X(x) = x \). Hence, there exists some \( k \in \mathbb{N} \) such that \( x = \sigma^k(x) \) (namely, \( k = 0 \)). In other words, \( x \sim x \) (by the definition of the relation \( \sim \)).

Now, forget that we fixed \( x \). We thus have shown that every \( x \in X \) satisfies \( x \sim x \). In other words, the relation \( \sim \) is reflexive.

20 Proof. Let \( x \in X \) and \( y \in X \) be such that \( x \sim y \). We shall show that \( y \sim x \).

Indeed, we have \( x \sim y \). In other words, there exists some \( k \in \mathbb{N} \) such that \( y = \sigma^k(x) \) (by the definition of the relation \( \sim \)). Consider such a \( k \), and denote it by \( u \). Thus, \( u \in \mathbb{N} \) satisfies \( y = \sigma^u(x) \).

Observation 1 yields that there exists some positive integer \( p \) such that \( \sigma^p(x) = x \). Consider this \( p \). Hence, Lemma 0.20 (applied to \( f = \sigma \) and \( k = u \)) yields \( \sigma^{up}(x) = x \). But \( p \) is positive; hence, \( p \geq 1 \) and thus \( up \geq u1 = u \). Hence, \( up - u \in \mathbb{N} \). Hence, \( \sigma^{up-u} \circ \sigma^u = \sigma^{(up-u)+u} = \sigma^{up} \). Thus, \( (\sigma^{up-u} \circ \sigma^u)(x) = \sigma^{up}(x) = x \). Hence, \( x = (\sigma^{up-u} \circ \sigma^u)(x) = \sigma^{up-u} \left( \sigma^u(x) \right) = \sigma^{up-u}(y) \).

Thus, there exists some \( k \in \mathbb{N} \) such that \( x = \sigma^k(y) \) (namely, \( k = up - u \)). In other words, \( y \sim x \) (by the definition of the relation \( \sim \)).

Now, forget that we fixed \( x \) and \( y \). We thus have shown that if \( x \in X \) and \( y \in X \) satisfy \( x \sim y \), then \( y \sim x \). In other words, the relation \( \sim \) is symmetric.

21 Proof. Let \( x, y, z \in X \) be such that \( x \sim y \) and \( y \sim z \). We shall show that \( x \sim z \).

Indeed, we have \( x \sim y \). In other words, there exists some \( k \in \mathbb{N} \) such that \( y = \sigma^k(x) \) (by the definition of the relation \( \sim \)). Consider such a \( k \), and denote it by \( u \). Thus, \( u \in \mathbb{N} \) satisfies \( y = \sigma^u(x) \).

Also, we have \( y \sim z \). In other words, there exists some \( k \in \mathbb{N} \) such that \( z = \sigma^k(y) \) (by the definition of the relation \( \sim \)). Consider such a \( k \), and denote it by \( v \). Thus, \( v \in \mathbb{N} \) satisfies \( z = \sigma^v(y) \).

But \( \sigma^p \circ \sigma^u = \sigma^{p+u} \). Thus, \( (\sigma^p \circ \sigma^u)(x) = \sigma^{p+u}(x) \). In view of \( (\sigma^p \circ \sigma^u)(x) = \sigma^p \left( \sigma^u(x) \right) \) = \( \sigma^p(y) = z \), this rewrites as \( z = \sigma^{p+u}(x) \). Thus, there exists some \( k \in \mathbb{N} \) such that \( z = \sigma^k(x) \) (namely, \( k = v + u \)). In other words, \( x \sim z \) (by the definition of the relation \( \sim \)).

Now, forget that we fixed \( x, y \) and \( z \). We thus have shown that if \( x \in X \), \( y \in X \) and \( z \in X \) satisfy \( x \sim y \) and \( y \sim z \), then \( x \sim z \). In other words, the relation \( \sim \) is transitive.
(b) Let \( x \in X \). Lemma 0.19(a) shows that there exists a \( j \in \mathbb{N} \) such that \( \sigma^j (x) \in \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{j-1} (x) \} \). Let \( p \) be the smallest such \( j \).

Lemma 0.19(b) shows that the integer \( p \) is positive and satisfies \( \sigma^p (x) = x \).

Lemma 0.19(c) shows that the elements \( \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \) are pairwise distinct. Lemma 0.19(d) shows that \( \{ \sigma^h (x) \mid h \in \mathbb{N} \} = \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \} \).

Define a set \( S \) by \( S = \{ \sigma^h (x) \mid h \in \mathbb{N} \} \). Thus,

\[
S = \{ \sigma^h (x) \mid h \in \mathbb{N} \} = \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \}.
\]

The definition of the equivalence class \([x]_\sim\) of \( x \) shows that

\[
[x]_\sim = \{ y \in X \mid y \sim x \}.
\]

Now, \([x]_\sim \subseteq S\) and \( S \subseteq [x]_\sim\). Combining these two relations, we obtain \([x]_\sim = S = \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \}\).

The \( p \) elements \( \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \) are pairwise distinct (as we have seen above). Thus, \(| \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \} | = p \).

Let \( k = |[x]_\sim|\). Then,

\[
k = |[x]_\sim| = \left| \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \} \right| = p.
\]

Now,

\[
[x]_\sim = \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{p-1} (x) \} = \{ \sigma^0 (x), \sigma^1 (x), \ldots, \sigma^{k-1} (x) \}
\]

(since \( p = k \)). This solves Exercise 7(b).

(c) Let \( E \) be a \( \sim \)-equivalence class. We must prove that \( c_E \) is a permutation of \( X \).

It is clear that \( c_E \) is well-defined. Next, we claim that

\[
t \in c_E (X) \quad \text{for each } t \in X.
\]

22Proof. Let \( w \in [x]_\sim \). Thus, \( w \in [x]_\sim = \{ y \in X \mid y \sim x \} \). In other words, \( w \) is an element of \( X \) and satisfies \( w \sim x \).

We have \( w \sim x \). Hence, \( x \sim w \) (since the relation \( \sim \) is symmetric). In other words, there exists some \( k \in \mathbb{N} \) such that \( w = \sigma^k (x) \) (by the definition of the relation \( \sim \)). Consider this \( k \). We have \( w = \sigma^k (x) \in \{ \sigma^h (x) \mid h \in \mathbb{N} \} = S \).

Now, forget that we fixed \( w \). We thus have shown that \( w \in S \) for each \( w \in [x]_\sim \). In other words, \([x]_\sim \subseteq S\).

23Proof. Let \( w \in S \). Thus, \( w \in S = \{ \sigma^h (x) \mid h \in \mathbb{N} \} \). In other words, \( w = \sigma^h (x) \) for some \( h \in \mathbb{N} \).

Consider this \( h \).

There exists some \( k \in \mathbb{N} \) such that \( w = \sigma^k (x) \) (namely, \( k = h \)). In other words, \( x \sim w \) (by the definition of the relation \( \sim \)). Thus, \( w \sim x \) (since the relation \( \sim \) is symmetric).

Hence, \( w \) is an element of \( X \) and satisfies \( w \sim x \). In other words, \( w \in [y \in X \mid y \sim x \} \). In view of \( [x]_\sim = \{ y \in X \mid y \sim x \} \), this rewrites as \( w \in [x]_\sim \).

Now, forget that we fixed \( w \). We thus have shown that \( w \in [x]_\sim \) for each \( w \in S \). In other words, \( S \subseteq [x]_\sim \).
[Proof of (35): Let \( t \in X \). We must prove that \( t \in c_E(X) \).

We are in one of the following two cases:

**Case 1:** We have \( t \in E \).

**Case 2:** We have \( t \notin E \).

Let us first consider Case 1. In this case, we have \( t \in E \). But \( E \) is an \( \sim \)-equivalence class. Hence, \( E \) is an \( \sim \)-equivalence class containing \( t \) (since \( t \in E \)). In other words, \( E = [t]_\sim \) (since the only \( \sim \)-equivalence class containing \( t \) is \( [t]_\sim \)).

Recall that \( \sigma \) is a permutation of \( X \). Hence, an element \( \sigma^{-1}(t) \) of \( X \) is well-defined. Denote this element by \( z \). Thus, \( z = \sigma^{-1}(t) \). Hence, \( \sigma(z) = t \).

We have \( \sigma^1(z) = \sigma(z) = t \), so that \( t = \sigma^1(z) \). Hence, there exists some \( k \in \mathbb{N} \) such that \( t = \sigma^k(z) \) (namely, \( k = 1 \)). In other words, \( z \sim t \) (by the definition of the relation \( \sim \)). Hence, \( z \) is an element of \( X \) and satisfies \( z \sim t \). In other words, \( z \in \{ y \in X \mid y \sim t \} \). But \( E = [t]_\sim = \{ y \in X \mid y \sim t \} \) (by the definition of the equivalence class \([t]_\sim \)). Hence, \( z \in \{ y \in X \mid y \sim t \} = E \).

The definition of \( c_E \) yields

\[
c_E(z) = \begin{cases} \sigma(z), & \text{if } z \in E; \\ z, & \text{if } z \notin E \end{cases} \quad \text{(since } z \in E \text{)}
\]

Hence, \( t = c_E \left( \frac{z}{z \in X} \right) \in c_E(X) \). Thus, we have proven \( t \in c_E(X) \) in Case 1.

Let us now consider Case 2. In this case, we have \( t \notin E \). The definition of \( c_E \) yields

\[
c_E(t) = \begin{cases} \sigma(t), & \text{if } t \in E; \\ t, & \text{if } t \notin E \end{cases} \quad \text{(since } t \notin E \text{)}.
\]

Hence, \( t = c_E \left( \frac{t}{t \in X} \right) \in c_E(X) \). Thus, we have proven \( t \in c_E(X) \) in Case 2.

We have now proven \( t \in c_E(X) \) in each of the two Cases 1 and 2. Hence, \( t \in c_E(X) \) is proven. This proves (35).

Now, (35) shows that \( X \subseteq c_E(X) \). In other words, the map \( c_E \) is surjective. Thus, \( c_E \) is a surjective map between two finite sets of the same size (namely, \( X \) and \( X \)), and therefore must be bijective (since any surjective map between two finite sets of the same size is bijective). In other words, \( c_E \) is a bijection \( X \rightarrow X \), therefore a permutation of \( X \). This solves Exercise 7(c).

(d) Let \( x \in X \) be such that \( E = [x]_\sim \). Let \( k = |[x]_\sim| \). We must prove that \( c_E \) can be written as \( \text{cyc}_{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)} \) (and in particular, we must prove that \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \) are distinct, so that \( \text{cyc}_{\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)} \) is well-defined).

Lemma 0.19(a) shows that there exists a \( j \in \mathbb{N} \) such that \( \sigma^j(x) \in \{ \sigma^0(x), \sigma^1(x), \ldots, \sigma^{j-1}(x) \} \). Let \( p \) be the smallest such \( j \). As in the solution to Exercise 7(b) (which we have given above), we can see that \( k = p \).

Lemma 0.19(b) shows that the integer \( p \) is positive and satisfies \( \sigma^p(x) = x \).

In view of \( k = p \), this rewrites as follows: The integer \( k \) is positive and satisfies \( \sigma^k(x) = x \). Since the integer \( k \) is positive, we have \( 1 \in [k] \).

Lemma 0.19(c) shows that the elements \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{p-1}(x) \) are pairwise distinct. In view of \( k = p \), this rewrites as follows: The elements \( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \) are pairwise distinct.
are pairwise distinct. Hence, the permutation \( \text{cyc}_{\sigma^0(x),\sigma^1(x),\ldots,\sigma^{k-1}(x)} \) is well-defined (since \( k \) is a positive integer).

Exercise 7(b) shows that \( [x]_\sim = \{ \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \} \). Hence,

\[
\left\{ \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \right\} = [x]_\sim = E
\]  

(since \( E = [x]_\sim \)).

It remains to prove that \( c_E \) can be written as \( \text{cyc}_{\sigma^0(x),\sigma^1(x),\ldots,\sigma^{k-1}(x)} \).

We define a \( k \)-tuple \( (i_1, i_2, \ldots, i_k) \) of elements of \( X \) by

\[
(i_1, i_2, \ldots, i_k) = \left( \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \right).
\]  

Thus,

\[
i_u = \sigma^{u-1}(x) \quad \text{for each } u \in [k].
\]  

Applying this to \( u = 1 \), we obtain \( i_1 = \sigma^{1-1}(x) \) (since \( 1 \in [k] \)).

Also, from (37), we obtain

\[
\{i_1, i_2, \ldots, i_k\} = \left\{ \sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x) \right\} = E
\]  

(by (36)).

We also let \( i_{k+1} \) stand for \( i_1 \). Thus, \( i_{k+1} = i_1 = \sigma^{1-1}(x) = \text{id}(x) = x = \sigma^k(x) \) (since \( \sigma^k(x) = x \)). Therefore, we see that

\[
i_u = \sigma^{u-1}(x) \quad \text{for each } u \in [k+1].
\]  

[Proof of (40): Let \( u \in [k+1] \). We must prove that \( i_u = \sigma^{u-1}(x) \). If \( u \in [k] \), then this follows from (38). Hence, for the rest of this proof, we WLOG assume that \( u \notin [k] \). Combining \( u \in [k+1] \) with \( u \notin [k] \), we obtain \( u \in [k+1] \setminus [k] = [k+1] \), so that \( u = k+1 \). Thus, \( i_u = i_{k+1} = \sigma^k(x) = \sigma^{u-1}(x) \) (since \( k = u - 1 \) (since \( u = k + 1 \))). This proves (40).]

We have \( \text{cyc}_{\sigma^0(x),\sigma^1(x),\ldots,\sigma^{k-1}(x)} = \text{cyc}_{i_1,i_2,\ldots,i_k} \) (since \( (\sigma^0(x), \sigma^1(x), \ldots, \sigma^{k-1}(x)) = (i_1, i_2, \ldots, i_k) \)). But the definition of \( \text{cyc}_{i_1,i_2,\ldots,i_k} \) yields

\[
\text{cyc}_{i_1,i_2,\ldots,i_k}(p) = \begin{cases} 
i_{j+1}, & \text{if } p = i_j \text{ for some } j \in \{1,2,\ldots,k\} \\
p, & \text{otherwise} \end{cases}
\]  

for every \( p \in X \).

Now, we claim that

\[
\text{cyc}_{i_1,i_2,\ldots,i_k}(p) = c_E(p) \quad \text{for each } p \in X.
\]  

[Proof of (42): Let \( p \in X \). We must prove the equality (42). We are in one of the following two cases:

Case 1: We have \( p = i_j \) for some \( j \in \{1,2,\ldots,k\} \).]
Case 2: We don’t have \( p = i_1 \) for some \( j \in \{1, 2, \ldots, k\} \).

Let us first consider Case 1. In this case, we have \( p = i_j \) for some \( j \in \{1, 2, \ldots, k\} \). Consider this \( j \).

Thus, (31) simplifies to \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = i_{j+1} \).

We have \( j \in \{1, 2, \ldots, k\} = [k] \). Hence, (38) (applied to \( u = j \)) yields \( i_j = \sigma^{j-1}(x) \). Hence, 
\[
\sigma^{j-1}(x) = i_j = p.
\]

But \( j \in \{1, 2, \ldots, k\} \), so that \( j + 1 \in \{2, 3, \ldots, k + 1\} \subseteq [k + 1] \). Hence, (40) (applied to \( u = j + 1 \)) yields
\[
i_{j+1} = \sigma^{(j+1)-1}(x) = \sigma^j(x) = \sigma \left( \frac{\sigma^{j-1}(x)}{\sigma^j(x) = p} \right) = \sigma(p).
\]

However,
\[
p = i_j \in \{i_1,i_2,\ldots,i_k\} \quad \text{(since \( j \in \{1, 2, \ldots, k\} \))}
\]
\[
e = E \quad \text{(by (39)).}
\]

The definition of \( c_E \) now shows that
\[
c_E(p) = \begin{cases} 
\sigma(p), & \text{if } p \in E; \\
n, & \text{if } p \notin E = \sigma(p) \quad \text{(since } p \in E) 
\end{cases}
\]
\[- \quad \text{by (43).}
\]

Comparing this with \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = i_{j+1} \), we obtain \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = c_E(p) \). Thus, (42) is proven in Case 1.

Let us now consider Case 2. In this case, we don’t have \( p = i_j \) for some \( j \in \{1, 2, \ldots, k\} \). Thus, (41) simplifies to \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = p \).

But we don’t have \( p = i_j \) for some \( j \in \{1, 2, \ldots, k\} \). In other words, \( p \notin \{i_1,i_2,\ldots,i_k\} \). In view of (39), this rewrites as \( p \notin E \). The definition of \( c_E \) now shows that
\[
c_E(p) = \begin{cases} 
\sigma(p), & \text{if } p \in E; \\
n, & \text{if } p \notin E = p \quad \text{(since } p \notin E) 
\end{cases}
\]

Comparing this with \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = p \), we obtain \( \text{cyc}_{i_1,i_2,\ldots,i_k}(p) = c_E(p) \). Thus, (42) is proven in Case 2.

We have now proven (42) in each of the two Cases 1 and 2. This completes the proof of (42).

The equality (42) shows that \( \text{cyc}_{i_1,i_2,\ldots,i_k} = c_E \) (since both \( \text{cyc}_{i_1,i_2,\ldots,i_k} \) and \( c_E \) are maps \( X \rightarrow X \)). Thus,
\[
c_E = \text{cyc}_{i_1,i_2,\ldots,i_k} = \text{cyc}_{\sigma^0(x),\sigma^1(x),\ldots,\sigma^{k-1}(x)}
\]
(by (37)). In other words, \( c_E \) can be written as \( \text{cyc}_{\sigma^0(x),\sigma^1(x),\ldots,\sigma^{k-1}(x)} \). This concludes the solution to Exercise (d).

(e) Recall that the \( \sim \)-equivalence classes form a set partition of the set \( X \) (in fact, this holds for the equivalence classes of any equivalence relation on \( X \)). Thus, each element of \( X \) belongs to exactly one \( \sim \)-equivalence class. Since \( E_1, E_2, \ldots, E_m \) are all the \( \sim \)-equivalence classes (listed without repetition), we can rewrite this fact as follows: Each element of \( X \) belongs to exactly one of the sets \( E_1, E_2, \ldots, E_m \). Thus, the sets \( E_1, E_2, \ldots, E_m \) are disjoint. In other words, if \( i \) and \( j \) are two distinct elements of \( [m] \), then
\[
E_i \cap E_j = \emptyset.
\]
Now, fix $x \in X$. Define $y \in X$ by $y = \sigma (x)$. We are going to show that
$$(c_{E_1} \circ c_{E_2} \circ \cdots \circ c_{E_m})(x) = y.$$  

The element $x$ of $X$ belongs to exactly one of the sets $E_1, E_2, \ldots, E_m$ (since each element of $X$ belongs to exactly one of the sets $E_1, E_2, \ldots, E_m$). In other words, there is exactly one $i \in [m]$ such that $x \in E_i$. Consider this $i$.

Hence, $i$ is the only element $j \in [m]$ such that $x \in E_j$. Therefore, every $j \in [m]$ distinct from $i$ must satisfy
$$x \not\in E_j.$$  

We have $x \sim y$, Thus, $y \in [x]_\sim$. But recall that $E_i$ is a $\sim$-equivalence class (since $E_1, E_2, \ldots, E_m$ are all the $\sim$-equivalence classes) and contains $x$ (since $x \in E_i$). Hence, $E_i$ is the $\sim$-equivalence class of $x$. In other words, $E_i = [x]_\sim$. Hence, $y \in [x]_\sim = E_i$. Hence, every $j \in [m]$ distinct from $i$ must satisfy
$$y \not\in E_j.$$  

Furthermore, $\sigma_{E_j}(y) = y$ for each $j \in [m]$ satisfying $j > i$. Also, $\sigma_{E_j}(y) = y$ for each $j \in [m]$ satisfying $j < i$. Therefore, Lemma 0.22 (applied to $g_j = \sigma_{E_j}$) shows that $(c_{E_1} \circ c_{E_2} \circ \cdots \circ c_{E_m})(x) = y = \sigma (x)$.

24Proof. We have $y = \sigma_{E_i}(x) = \sigma^{E_i}(x)$. Thus, there exists some $k \in \mathbb{N}$ such that $y = \sigma^k (x)$ (namely, $k = 1$). In other words, $x \sim y$ (since $x \sim y$ if and only if there exists some $k \in \mathbb{N}$ such that $y = \sigma^k (x)$).

Proof of (45): Fix $j \in [m]$ distinct from $i$. We must show that $y \not\in E_j$.

Assume the contrary. Thus, $y \in E_j$. Combining this with $y \in E_i$, we find $y \in E_i \cap E_j$. Therefore, the set $E_i \cap E_j$ is nonempty (namely, it contains $y$). But $j$ is distinct from $i$. Hence, (44) yields $E_i \cap E_j = \emptyset$. This contradicts the fact that the set $E_i \cap E_j$ is nonempty. This contradiction shows that our assumption was false, qed.

26Proof. The definition of $\sigma_{E_i}$ yields
$$\sigma_{E_i}(x) = \begin{cases} \sigma (x), & \text{if } x \in E_i; \\ x, & \text{if } x \not\in E_i \end{cases} = \sigma (x) \quad \text{(since } x \in E_i)$$

27Proof. Let $j \in [m]$ be such that $j > i$. Thus, $j$ is distinct from $i$ (since $j > i$). Hence, (45) shows that $x \not\in E_j$. Now, the definition of $\sigma_{E_j}$ yields
$$\sigma_{E_j}(x) = \begin{cases} \sigma (x), & \text{if } x \in E_j; \\ x, & \text{if } x \not\in E_j \end{cases} = x \quad \text{(since } x \not\in E_j).$$

Qed.

28Proof. Let $j \in [m]$ be such that $j < i$. Thus, $j$ is distinct from $i$ (since $j < i$). Hence, (46) shows that $y \not\in E_j$. Now, the definition of $\sigma_{E_j}$ yields
$$\sigma_{E_j}(y) = \begin{cases} \sigma (y), & \text{if } y \in E_j; \\ y, & \text{if } y \not\in E_j \end{cases} = y \quad \text{(since } y \not\in E_j).$$
Now, forget that we fixed \( x \). We thus have shown that \( (c_{E_1} \circ c_{E_2} \circ \cdots \circ c_{E_m}) (x) = \sigma (x) \) for each \( x \in X \). In other words, \( c_{E_1} \circ c_{E_2} \circ \cdots \circ c_{E_m} = \sigma \). This solves Exercise 7(e).

References

https://eudml.org/doc/121542


See also http://www.cip.ifi.lmu.de/~grinberg/algebra/HopfComb-sols.pdf for a version that gets updated.

\[ \square \]

Qed.