Classical Finite Transformation Semigroups

Algebra and Applications

Volume 9

Algebra and Applications aims to publish well written and carefully refereed monographs with up-to-date information about progress in all fields of algebra, its classical impact on commutative and noncommutative algebraic and differential geometry, K-theory and algebraic topology, as well as applications in related domains, such as number theory, homotopy and (co)homology theory, physics and discrete mathematics.

Particular emphasis will be put on state-of-the-art topics such as rings of differential operators, Lie algebras and super-algebras, group rings and algebras, C*algebras, Kac-Moody theory, arithmetic algebraic geometry, Hopf algebras and quantum groups, as well as their applications. In addition, Algebra and Applications will also publish monographs dedicated to computational aspects of these topics as well as algebraic and geometric methods in computer science.

Olexandr Ganyushkin • Volodymyr Mazorchuk

Classical Finite Transformation Semigroups

An Introduction

Springer

Olexandr Ganyushkin
Kyiv Taras Shevchenko University
Volodymyrska Street, 64
Kyiv
Ukraine
ganiyshk@univ.kiev.ua

Volodymyr Mazorchuk
Uppsala University
Dept. Mathematics
SE-751 06 Uppsala
Sweden
mazor@math.uu.se

ISBN: 978-1-84800-280-7 e-ISBN: 978-1-84800-281-4
DOI: 10.1007/978-1-84800-281-4

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008941019

Printed on acid-free paper

Springer Science+Business Media
springer.com

Preface

Semigroup theory is a relatively young part of mathematics. As a separate direction of algebra with its own objects, formulations of problems, and methods of investigations, semigroup theory was formed about 60 years ago.

One of the main motivations for the existence of some mathematical theories are interesting and natural examples. For semigroup theory the obvious candidates for such examples are transformation semigroups. Various transformations of different sets appear everywhere in mathematics all the time. As the usual composition of transformations is associative, each set of transformations, closed with respect to the composition, forms a semigroup.

Among all transformation semigroups one can distinguish three classical series of semigroups: the *full symmetric semigroup* $\mathcal{T}(M)$ of all transformations of the set M; the *symmetric inverse semigroup* $\mathcal{IS}(M)$ of all partial (that is, not necessarily everywhere defined) injective transformations of M; and, finally, the semigroup $\mathcal{PT}(M)$ of all partial transformations of M. If $M = \{1, 2, \dots, n\}$, then the above semigroups are usually denoted by $\mathcal{T}_n$, $\mathcal{IS}_n$ and $\mathcal{PT}_n$, respectively. One of the main evidences for the importance of these semigroups is their universality property: every (finite) semigroup is a subsemigroup of some $\mathcal{T}(M)$ (resp. $\mathcal{T}_n$); and every (finite) inverse semigroup is a subsemigroup of some $\mathcal{IS}(M)$ (resp. $\mathcal{IS}_n$). Inverse semigroups form a class of semigroups which are closest (in some sense) to groups.

An analogous universal object in group theory is the *symmetric group* $\mathcal{S}(M)$ of all bijective transformations of M. Many books are dedicated to the study of $\mathcal{S}(M)$ or to the study of transformation groups in general. Transformation semigroups had much less luck. The "naive" search in MathSciNet for books with the keywords "transformation semigroups" in the title results in two titles, one being a conference proceedings, and another one being old 50-page-long lecture notes in Russian ([Sc4]). Just for comparison, an analogous search for "transformation groups" results in 75 titles. And this is in spite of the fact that the semigroup $\mathcal{T}(M)$ was studied by Suschkewitsch already in the 1930s. The semigroup $\mathcal{IS}(M)$ was introduced by Wagner in 1952, but the first relatively small monograph about it appeared only in 1996 ([Li]). The latter monograph considers some basic questions about $\mathcal{IS}(M)$: how one writes down the elements of $\mathcal{IS}(M)$, when two elements of $\mathcal{IS}(M)$ commute, what is the presentation of $\mathcal{IS}(M)$, what are the congruences on $\mathcal{IS}(M)$. For example, such basic semigroup-theoretical notions as ideals and Green's relations are mentioned only in the Appendix without any direct relation to $\mathcal{IS}(M)$.

Much more information about the semigroup $\mathcal{T}(M)$ can be found in the last chapter of [Hi1]; however, it is mostly concentrated around the combinatorial aspects. Otherwise one is left with the options to search through examples in the parts of the abstract theory of semigroups using the classical books [CP1, Gri, Ho3, Ho7, Hi1, Law, Pe] or to look at original research papers.

The aim of the present book is to partially fill the gaps in the literature. In the book we introduce three classical series of semigroups, and for them we describe generating systems, ideals, Green's relations, various classes of subsemigroups, congruences, conjugations, endomorphisms, presentations, actions on sets, linear representations and cross-sections. Some of the results are very old and classical, some are quite young. In order not to overload the reader with too technical and specialized results, we decided to restrict the area of the present book to the above-mentioned parts of the theory of transformation semigroups.

The book was thought to be an elementary introduction to the theory of transformation semigroups, with a strong emphasis on the concrete examples in the form of three classical series of finite transformation semigroups, namely, $\mathcal{T}_n$, $\mathcal{IS}_n$ and $\mathcal{PT}_n$. The book is primarily directed to students, who would like to make their first steps in semigroup theory. The choice of the semigroups $\mathcal{T}_n$, $\mathcal{IS}_n$ and $\mathcal{PT}_n$ is motivated not only by their role in semigroup theory, but also by our strong belief that a good understanding of a couple of interesting and pithy examples is more important for the first acquaintance with some theory than a formal learning of dozens of theorems.

Another good motivation to consider the semigroups $\mathcal{T}_n, \mathcal{IS}_n$ and $\mathcal{PT}_n$ at the same time is the observation that many results about these semigroups, which for each of them were obtained independently by different people and in different times, in reality can be obtained in a unified (or almost unified) way.

Several results which will be presented extend in one or another way to the cases of infinite transformation semigroups. However, we restrict ourselves to the case of finite semigroups to make the exposition as elementary and accessible for a wide audience as possible. We are not after the biggest possible generality. Another argument is that we also try to attract the reader's attention to numerous combinatorial aspects and applications of the semigroups we consider.

With our three principal examples of semigroups on the background we also would like to introduce the reader to the basics of the abstract theory of semigroups. So, along the discussion of these examples, we tried to present many important basic notions and prove (or at least mention) as many classical abstract results as possible.

The requirements for the reader's mathematical background are very low. To understand most of the content, it is enough to have a minimal mathematical experience on the level of common sense. Perhaps some familiarity with basic university courses in algebra and combinatorics would be a substantial help. We have tried to define all the notions we use in the book. We have also tried to make all proofs very detailed and to avoid complicated constructions whenever possible.

The penultimate section of each chapter is called "Addenda and Comments." A part of it consists of historical comments (which are by no means complete). Another part consists of some remarks, facts, and statements, which we did not include in the main text of the book. The reason is usually the much less elementary level of these statements or the more complicated character of the proofs. However, we include them in the Addenda as from our point of view they deserve attention in spite of the fact that they do not really fit into the main text. Some statements here are also given with proofs, but these proofs are less detailed than those in the main text. For this part of the book, our requirements for the reader's mathematical background are different and are closer to the standard mathematical university curriculum. In the Addenda, we sometimes also mention some open problems and try to describe possible directions for further investigations.

The division of the book into the main text and the Addenda is not very strict as sometimes the notions and facts mentioned in the Addenda are used in the main text.

The last section of each chapter contains problems. Some problems (not many) are also included in the main text. The latter ones are mostly simple and directed to the reader. Sometimes they also ask to repeat a proof given before for a different situation. These problems are in some sense compulsory for the successful understanding of the main text (i.e., one should at least read them). The additional problems of the last section of each chapter can be quite different. Some of them are easy exercises, while others are much more complicated problems, which form an essential supplement to the material of the chapter. Hints for solutions of the latter ones can be found at the end of the book.

The book was essentially written during the visit of the first author to Uppsala University, which was supported by The Royal Swedish Academy of Sciences and The Swedish Foundation for International Cooperation in Research and Higher Education (STINT). The financial support of The Academy and STINT, and the hospitality of Uppsala University are gratefully acknowledged. We thank Ganna Kudryavtseva, Victor Maltcev, and Abdullahi Umar for their comments on the preliminary version of the book.

Kyiv, Ukraine — *Olexandr Ganyushkin*
Uppsala, Sweden — *Volodymyr Mazorchuk*

Contents

Chapter 1

Ordinary and Partial Transformations

1.1 Basic Definitions

The principal objects of interest in the present volume are finite sets and transformations of finite sets. Let M be a finite set, say $M=\{m_1, m_2, \dots, m_n\}$, where n is a nonnegative integer. *Transformation* of M is an array of the following form:

$$\alpha = \begin{pmatrix} m_1 & m_2 & \cdots & m_n \\ k_1 & k_2 & \cdots & k_n \end{pmatrix}, \tag{1.1}$$

where all $k_i \in M$. If $x \in M$, say $x = m_i$, the element k_i will be called the *value* of the transformation α at the element x and will be denoted by $\alpha(x)$. The fact that α is a transformation of M is usually written as $\alpha : M \to M$. As the nature of elements of M is not important for us, instead of M we shall usually consider the set $\mathbf{N} = \mathbf{N}_n = \{1, 2, \dots, n\}$.

Apart from the transformations of M we shall also consider the so-called *partial transformations* of M, that is, transformations of the form $\alpha : A \to M$, where $A = \{l_1, l_2, \dots, l_k\}$ is a subset of M. Note that the set A can be empty. Again, the element α can be written in the following *tabular form*:

$$\alpha = \begin{pmatrix} l_1 & l_2 & \cdots & l_k \\ \alpha(l_1) & \alpha(l_2) & \cdots & \alpha(l_k) \end{pmatrix}. \tag{1.2}$$

Abusing notation, we may also write $\alpha : M \to M$ for a partial transformation, having in mind that such α is only defined on some elements from M. Note that the order of elements in the first row of arrays (1.1) and (1.2) is not important.

With each (partial) transformation α as above we associate the following standard notions:

- The *domain* of α: $\mathrm{dom}(\alpha) = A$

O. Ganyushkin, V. Mazorchuk, *Classical Finite Transformation Semigroups,* Algebra and Applications 9, DOI: 10.1007/978-1-84800-281-4_1,

- The *codomain* of α: $\overline{\mathrm{dom}}(\alpha) = M \backslash A$
- The *image* of α: $\mathrm{im}(\alpha) = \{\alpha(x) \,:\, x \in A\}$

The word *range*, which is also frequently used in the literature, is a synonym of the word *image*. If $\mathrm{dom}(\alpha) = M$, the transformation α is called *full* or *total* .

The set of all total transformations of M is denoted by $\mathcal{T}(M)$, and the set of all partial transformations of M is denoted by $\mathcal{PT}(M)$. Obviously, $\mathcal{T}(M) \subset \mathcal{PT}(M)$. To simplify our notation we set $\mathcal{T}_n = \mathcal{T}(\mathbf{N})$ and $\mathcal{PT}_n = \mathcal{PT}(\mathbf{N})$.

Sometimes it is convenient to use a slightly modified version of (1.2) for some $\alpha \in \mathcal{PT}_n$. In the case of $\mathcal{PT}_n$ it is natural to form the first row of the array for α by simply listing all the elements from $\mathbf{N}$ in their natural order. Then, to define α completely, one needs a special symbol to indicate that some element x belongs to $\overline{\mathrm{dom}}(\alpha)$. We shall use the symbol $\varnothing$. In other words, $\alpha(x) = \varnothing$ means that $x \in \overline{\mathrm{dom}}(\alpha)$. Thus the element α can be written in the following form:

$$\alpha = \begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix}, \tag{1.3}$$

where $k_i = \alpha(i)$ if $i \in \mathrm{dom}(\alpha)$ and $k_i = \varnothing$ if $i \in \overline{\mathrm{dom}}(\alpha)$.

Example 1.1.1 Here is the list of all elements of $\mathcal{PT}_2$:

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 \\ 1 & \varnothing \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & \varnothing \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ \varnothing & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ \varnothing & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ \varnothing & \varnothing \end{pmatrix}.$$

The first row of this list consists of total transformations and hence lists all elements in $\mathcal{T}_2$.

Proposition 1.1.2 *The set $\mathcal{T}_n$ contains n^n elements and the set $\mathcal{PT}_n$ contains $(n+1)^n$ elements.*

Proof. Each element $\alpha \in \mathcal{T}_n$ is uniquely defined by array (1.3), where each $k_i \in \mathbf{N}$. Since the choices of k_is are independent, we have $|\mathcal{T}_n| = n^n$ by the product rule. In the case of $\mathcal{PT}_n$, the elements k_i can be independently chosen from the set $\mathbf{N} \cup \{\varnothing\}$. Hence the product rule implies $|\mathcal{PT}_n| = (n+1)^n$. □

The cardinality $|\mathrm{im}(\alpha)|$ of the image of a partial transformation $\alpha \in \mathcal{PT}_n$ is called the *rank* of this partial transformation and is denoted by $\mathrm{rank}(\alpha)$. Thus $\mathrm{rank}(\alpha)$ equals the number of different elements in the second row of array (1.2). The number $\mathrm{def}(\alpha) = n - \mathrm{rank}(\alpha)$ is called the *defect* of the partial transformation α.

A partial transformation $\alpha \in \mathcal{PT}_n$ is called

- *Surjective* if $\mathrm{im}(\alpha) = \mathbf{N}$
- *Injective* if $x \neq y$ implies $\alpha(x) \neq \alpha(y)$ for all $x, y \in \mathrm{dom}(\alpha)$
- *Bijective* if α is both surjective and injective

If α is given by (1.2), then surjectivity means that the second row of array (1.2) contains all elements of $\mathbf{N}$; and injectivity means that all elements in the second row of array (1.2) are different. Bijective transformations on $\mathbf{N}$ are also called *permutations* of $\mathbf{N}$.

Proposition 1.1.3 *Let $\alpha \in \mathcal{T}_n$. Then the following conditions are equivalent:*

(a) α is surjective

(b) α is injective

(c) α is bijective

Proof. By the definition of a bijective transformation, it is enough to show that the conditions (a) and (b) are equivalent. We start by proving that injectivity implies surjectivity. Let $\alpha \in \mathcal{T}_n$ be injective and given by (1.3). Because of the injectivity of α, the second row of (1.3) gives n different elements of the set $\mathbf{N}$, namely, $\alpha(1), \alpha(2), \dots, \alpha(n)$. But $\mathbf{N}$ contains exactly n elements. Hence $\mathbf{N} = \{\alpha(1), \alpha(2), \dots, \alpha(n)\}$, and thus α is surjective.

Conversely, let $\alpha \in \mathcal{T}_n$ be surjective and given by (1.3). Then the second row of (1.3) contains all n elements of the set $\mathbf{N}$. But this row contains exactly n elements. Hence they all must be different. This implies that α is injective. $\square$

1.2 Graph of a (Partial) Transformation

With each partial transformation α on $\mathbf{N}$ one naturally associates a directed graph Γ_α. A *directed graph* (or a *digraph*) is a pair, $\Gamma = (V, E)$, where V is a set and $E \subset V \times V$. The elements of V are called *vertices* of Γ and the elements of E are called *directed edges* or *arrows* of Γ. If $(a, b) \in E$, then the vertex a is called the *tail* of (a, b) and the vertex b is called the *head* of (a, b).

The graph $\Gamma_\alpha = (V_\alpha, E_\alpha)$ is called the *graph of the transformation* α and is constructed in the following way: The set V_α of vertices coincides with $\mathbf{N}$; for $x, y \in \mathbf{N}$ the element (x, y) belongs to E_α if and only if $x \in \mathrm{dom}(\alpha)$ and $\alpha(x) = y$.

Example 1.2.1 For the transformation

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 8 & 9 & 7 & 13 & 7 & 16 & \varnothing & 1 & 13 & 11 & 16 & 4 & 12 & 9 & 14 & 13 \end{pmatrix}$$

the graph Γ_α has the following form:

$$\begin{tikzcd}
10 \arrow[r] & 11 \arrow[dr] & & 6 \arrow[dl] & 8 \arrow[d, bend left] & 3 \arrow[d] & 5 \arrow[dl] \\
15 \arrow[d] & 2 \arrow[d] & 16 \arrow[d] & 4 \arrow[dl] & 1 \arrow[u, bend left] & 7 & \\
14 \arrow[r] & 9 \arrow[r] & 13 \arrow[r] & 12 \arrow[u] & & &
\end{tikzcd} \tag{1.4}$$

It is obvious that a directed graph $\Gamma = (\mathbf{N}, E)$ will be the graph of some total (partial) transformation of $\mathbf{N}$ if and only if each vertex is the tail of exactly one (at most one) arrow. The graph Γ_α decomposes into a disjoint union of connected components. Intuitively, this is an obvious notion, for example, graph (1.4) has three connected components. The rigorous definition is as follows.

First we define a subgraph. If $\Gamma = (V, E)$ is a directed graph, a *subgraph* of Γ is a directed graph $\Gamma' = (V', E')$ such that $V' \subset V$ and $E' \subset E$. A directed graph $\Gamma = (V, E)$ is called *connected* if for each partition of V into a disjoint union of nonempty subsets V_1 and V_2 there exists $a \in V_1$ and $b \in V_2$ such that either (a, b), or (b, a) is an arrow. If Γ is a directed graph, then *the connected components* of Γ are simply the *maximal connected* subgraphs of Γ, that is, those connected subgraphs of Γ which are not proper subgraphs of any other connected subgraph of Γ.

Exercise 1.2.2 Prove that two different connected components of a directed graph Γ do not have common arrows.

To understand the structure of Γ_α it is of course enough to understand the structure of its connected components. For this we shall need some more graph-theoretical notions. Let $\Gamma(V, E)$ be a directed graph and $a, b \in V$. An *oriented path* from a to b in Γ is a sequence $x_0 = a, x_1, \dots, x_k = b$ of vertices such that $(x_i, x_{i+1}) \in E$ for each $i = 0, 1, \dots, k-1$. Vertex a is called the *tail* of the path and vertex b is called the *head* of the path. If we have an oriented path such that $a = b$ and $x_i \neq x_j$ for all $0 \leq i < j < k$, then such path is called an (*oriented*) *cycle* and is denoted by $(x_0, x_1, \dots, x_{k-1})$. If Γ does not contain any arrow with tail b we will say that our path *breaks* at b. Analogously one defines infinite paths. Such paths may be without tails, without heads, or without both tails and heads.

Let Γ be a directed graph and v be a vertex of Γ. We define a *trajectory* of v as any longest possible path with the tail v. There are two possibilities:

either such trajectory is finite and hence breaks at some point, or this trajectory is infinite. For example, 5, 7 is a trajectory of vertex 5 in graph (1.4), which breaks at vertex 7; and 2, 9, 13, 12, 4, 13, 12, 4, 13, $\dots$ is a trajectory of point 2. In general, a point can have many different trajectories. However, we have the following obvious statement.

Lemma 1.2.3 *Let $\varGamma$ be a directed graph. Then the following conditions are equivalent:*

(a) Each vertex of $\varGamma$ has a unique trajectory

(b) Each vertex of $\varGamma$ is the tail of at most one arrow

We will say that the infinite trajectory $x_0 = v$, $x_1, \dots$ *terminates* at the cycle $(x_k, x_{k+1}, \dots, x_{k+m-1})$ if the path $x_k, x_{k+1}, \dots, x_{k+m-1}$ is a cycle, $x_i = x_{i+m}$ for all $i \geq k$ and $x_{k-1} \neq x_{k+m-1}$. Thus, the trajectory of vertex 2 in graph (1.4) terminates at the cycle $(13, 12, 4)$; and the trajectory of vertex 4 terminates at the cycle $(4, 13, 12)$.

Proposition 1.2.4 *Let $\alpha \in \mathcal{PT}_n$.*

(i) Every vertex of $\varGamma_\alpha$ has a unique trajectory.

(ii) The trajectory of each $x \in \mathbf{N}$ in $\varGamma_\alpha$ either breaks at some vertex or terminates at some cycle.

(iii) α is total if and only if the trajectory of each $x \in \mathbf{N}$ in $\varGamma_\alpha$ terminates at some cycle.

(iv) Let $x, y \in \mathbf{N}$. If y occurs in the trajectory of x in $\varGamma_\alpha$, then the trajectory of y is a subsequence of the trajectory of x.

Proof. Since each vertex of $\varGamma_\alpha$ is a tail of at most one arrow, the statement (i) follows immediately from Lemma 1.2.3. The statement (iv) follows immediately from (i).

Assume that the trajectory $x = x_0$, $x_1, \dots$ of x does not break. Since $\varGamma_\alpha$ is finite, this trajectory must contain repetitions of some vertices. Let k be minimal for which there exists a repetition of x_k and let x_{k+m} be the first repetition of x_k. Since each vertex of $\varGamma_\alpha$ is a tail of at most one arrow, the condition $x_k = x_{k+m}$ implies $x_{k+1} = x_{k+m+1}$, which, in turn, implies $x_{k+2} = x_{k+m+2}$, and so on. Hence our trajectory terminates at the cycle $(x_k, x_{k+1}, \dots, x_{k+m-1})$. This proves (ii).

If α is total, the trajectory of each vertex cannot break. Hence (iii) follows from (ii). $\square$

For $\alpha \in \mathcal{PT}_n$ and $x \in \mathbf{N}$ we denote by $\mathrm{tr}_\alpha(x)$ the trajectory of x in $\varGamma_\alpha$. This is well-defined because of Proposition 1.2.4(i). Define now the binary relation ω_α on $\mathbf{N}$ in the following way: For $x, y \in \mathbf{N}$ set $x\,\omega_\alpha\,y$ if $\mathrm{tr}_\alpha(x)$ and $\mathrm{tr}_\alpha(y)$ have at least one common vertex.

Lemma 1.2.5 *The relation ω_α is an equivalence relation.*

Proof. That ω_α is reflexive and symmetric is obvious. To prove the transitivity of ω_α consider $x, y, z \in \mathbf{N}$ such that $x\,\omega_\alpha\,y$ and $y\,\omega_\alpha\,z$. Let a be a common vertex of $\mathrm{tr}_\alpha(x)$ and $\mathrm{tr}_\alpha(y)$ and b be a common vertex of $\mathrm{tr}_\alpha(y)$ and $\mathrm{tr}_\alpha(z)$. Without loss of generality we can assume that the first occurrence of a in $\mathrm{tr}_\alpha(y)$ is not later than the first occurrence of b in $\mathrm{tr}_\alpha(y)$. But this means that b occurs in $\mathrm{tr}_\alpha(a)$ by Proposition 1.2.4(iv). Another application of Proposition 1.2.4(iv) implies that b occurs in $\mathrm{tr}_\alpha(x)$. Hence $x\,\omega_\alpha\,z$, completing the proof. $\square$

The equivalence classes of ω_α are called the *orbits* of α. For $x \in \mathbf{N}$ the orbit of x in Γ_α will be denoted by $\mathfrak{o}_\alpha(x)$. From the definition of ω_α it follows that for any $x \in \mathrm{dom}(\alpha)$ we have $x\,\omega_\alpha\,\alpha(x)$. Hence all vertices which occur in $\mathrm{tr}_\alpha(x)$ belong to $\mathfrak{o}_\alpha(x)$. Furthermore, for each $x \in \mathbf{N}$ we can restrict the partial transformation α to the orbit $K = \mathfrak{o}_\alpha(x)$, obtaining a new partial transformation, $\alpha^{(K)} \in \mathcal{PT}(\mathfrak{o}_\alpha(x))$. Certainly, $\alpha^{(K)}$ does not depend on the choice of the vertex in K.

Proposition 1.2.6 *For each $x \in \mathbf{N}$ the graph $\Gamma_{\alpha^{(K)}}$ is a connected component of Γ_α.*

Proof. From the definition of ω_α it follows that the graph $\Gamma_{\alpha^{(K)}}$ is connected and contains all those arrows of Γ_α, for which both the heads and the tails belong to K. Assume now that (x, y) is an arrow of Γ_α such that $x \in K$. Then $y \in \mathrm{tr}_\alpha(x)$ and hence $x\,\omega_\alpha\,y$, that is, $y \in K$. If (x, y) is an arrow of Γ_α such that $y \in K$, then again $y \in \mathrm{tr}_\alpha(x)$ and hence $x\,\omega_\alpha\,y$, that is, $x \in K$. This means that $\Gamma_{\alpha^{(K)}}$ is not properly contained in any connected subgraph of Γ_α, which proves our statement. $\square$

As an immediate corollary of Proposition 1.2.6 we have:

Corollary 1.2.7 *The mapping $K \mapsto \Gamma_{\alpha^{(K)}}$ is a bijection between the orbits of α and the connected components of Γ_α.*

A directed graph $\Gamma = \Gamma(V, E)$ is called a *tree* with the *sink* $a \in V$ provided that for each $x \in V$ the trajectory of x in Γ is unique and breaks at a. For instance, in the example (1.4) if $K = \mathfrak{o}_\alpha(3)$, the connected component $\Gamma_{\alpha^{(K)}}$ is a tree with the sink 7. A (nonempty) disjoint union of several trees with sinks is called a *forest* of trees with sinks.

Exercise 1.2.8 Let Γ be a tree with the sink a. Show that Γ is connected; that each $b \neq a$ is a tail of exactly one arrow; and that Γ contains neither oriented nor unoriented cycles.

A directed graph $\Gamma = (V, E)$ is called a *cycle* provided that we can enumerate $V = \{a_1, \dots, a_k\}$ such that $E = \{(a_1, a_2), (a_2, a_3), \dots, (a_{k-1}, a_k),$

$(a_k, a_1)\}$. If $\Gamma_i = (V_i, E_i)$, $i \in I$, are directed graphs, then their *union* $\Gamma = \cup_{i\in I}\Gamma_i$ is defined as follows $\Gamma = (V, E)$, where $V = \cup_{i\in I} V_i$ and $E = \cup_{i\in I} E_i$. For example, each directed graph is a union of its connected components.

Theorem 1.2.9 *Each connected component of Γ_α, $\alpha \in \mathcal{PT}_n$, is either*

(i) a tree with a sink, or

(ii) a union of a forest of trees with sinks with a cycle on the set of all their sinks.

We note that the forest in Theorem 1.2.9(ii) may contain only one tree with a sink. The union of this tree with a sink with the cycle on its sink is not a tree with a sink anymore. We also note that the forest in Theorem 1.2.9(ii) may also have some *trivial* trees with sinks, that is, trees consisting only of sinks. If all trees in this forest are trivial, Theorem 1.2.9(ii) simply describes a cycle.

Proof. Let $\alpha \in \mathcal{PT}_n$ and $\Gamma_{\alpha(K)} = (K, E_K)$ be a connected component of Γ_α, and $x \in K$. Then Lemma 1.2.3 and Proposition 1.2.4(i) imply that x has a unique trajectory in both Γ_α and $\Gamma_{\alpha(K)}$. Moreover, since K is a connected component of Γ_α we also have that the trajectories of x in Γ_α and $\Gamma_{\alpha(K)}$ coincide (and hence they both are equal to $\mathrm{tr}_\alpha(x)$).

Assume first that $\mathrm{tr}_\alpha(x)$ breaks at some vertex, say a. Let $y \in K$ be arbitrary. Then, by definition, $\mathrm{tr}_\alpha(x)$ and $\mathrm{tr}_\alpha(y)$ have a common vertex, say z. Hence $\mathrm{tr}_\alpha(z)$ is a subsequence of both $\mathrm{tr}_\alpha(x)$ and $\mathrm{tr}_\alpha(y)$. But $\mathrm{tr}_\alpha(x)$ breaks at a, which means that $\mathrm{tr}_\alpha(z)$ must break at a as well. This implies that $\mathrm{tr}_\alpha(y)$ breaks at a. This means, by definition, that $\Gamma_{\alpha(K)}$ is a tree with the sink a, that is, of the type Theorem 1.2.9(i).

Now we assume that $\mathrm{tr}_\alpha(x)$ terminates at some cycle, say $(a_1, a_2, \dots, a_k)$. For each a_i, $i = 1, \dots, k$, the trajectory of a_i is unique by Proposition 1.2.4(i) and hence is $a_i, a_{i+1}, \dots, a_k, a_1, \dots$. For every $y \in K$, the trajectory $\mathrm{tr}_\alpha(y)$ has a common subsequence with $\mathrm{tr}_\alpha(x)$ and thus must contain some a_i. For $i = 1, \dots, k$ we denote by K_i the set of all those vertices y from K such that the first vertex from the cycle $(a_1, a_2, \dots, a_k)$ in $\mathrm{tr}_\alpha(y)$ is a_i. Note that $K_i \cap K_j = \varnothing$ for $i \neq j$ by definition.

Assume that $i \neq j$ and let (x, y) be an arrow in $\Gamma_{\alpha(K)}$ such that $x \in K_i$ and $y \in K_j$. Then $\mathrm{tr}_\alpha(x)$ has the form $x, y, y_1, \dots,$ where $y, y_1, \dots$ is just $\mathrm{tr}_\alpha(y)$. However, the first element from the cycle $(a_1, a_2, \dots, a_k)$ in $\mathrm{tr}_\alpha(x)$ is a_i, whereas in $\mathrm{tr}_\alpha(y)$ it is $a_j \neq a_i$. This is possible only in the case of $x = a_i$ and $y = a_j$. This means that every arrow in $\Gamma_{\alpha(K)}$ from some element in K_i to some element in K_j in fact belongs to the cycle $(a_1, a_2, \dots, a_k)$.

For each $i = 1, \dots, k$, consider the graph $\Gamma_i = (K_i, E_i)$, where

$$E_i = ((K_i \times K_i) \cap E) \backslash \{(a_i, a_i)\},$$

(this means that E_i consists of all arrows from E, which has both tails and heads in K_i, with the exception of the arrow (a_i, a_i)). Let Γ_0 be the cycle $(a_1, a_2, \dots, a_k)$. From the previous paragraph, we have that $\Gamma_{\alpha(K)} = \cup_{i=0}^{k} \Gamma_i$. Furthermore, K_is are disjoint for $i = 1, \dots, k$. To show that $\Gamma_{\alpha(K)}$ is of the form described in Theorem 1.2.9(ii) it remains to show that each Γ_i, $i = 1, \dots, k$, is a tree with the sink a_i.

By definition, Γ_i does not contain any arrow with the tail a_i. Let $y \in K_i$. Then $\mathrm{tr}_\alpha(y)$ has the form $y = y_0, y_1, \dots, y_m = a_i, a_{i+1}, \dots$, where $y_m = a_i$ is the first occurrence of a_i in $\mathrm{tr}_\alpha(y)$. By the definition of K_i, all vertices $y_1, \dots, y_{m-1}$ do not belong to Γ_0. Hence the definition of E_i implies that $y = y_0, y_1, \dots, y_m = a_i$ is the trajectory of y in Γ_i, and it breaks at a_i. In other words, the trajectory of each vertex in Γ_i breaks at a_i and thus Γ_i is a tree with the sink a_i. This completes the proof. □

Example 1.2.10 The graph Γ_α from Example 1.2.1 is given by (1.4) and has three connected components. The third component is a tree with the sink 7. The second component is a cycle (that is, the union of the cycle $(1, 8)$ with the corresponding forest of trivial trees with sinks). The first component is the union of the cycle $(4, 13, 12)$ with the following forest of trees with sinks:

$$
\begin{tikzcd}
15 \arrow[r] & 14 \arrow[dr] & & & & 11 \arrow[dl] & 10 \arrow[l] & 12 & 4 \\
& 2 \arrow[r] & 9 \arrow[dr] & & 16 \arrow[dl] & 6 \arrow[l] & & & \\
& & & 13 & & & & &
\end{tikzcd}
$$

An immediate corollary of Theorem 1.2.9 is the following:

Corollary 1.2.11 *Different cycles of Γ_α belong to different connected components of Γ_α.*

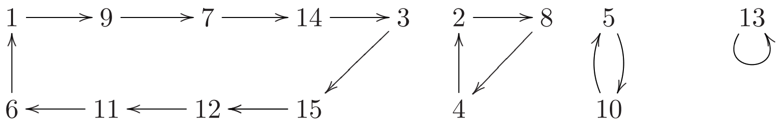
1.3 Linear Notation for Partial Transformations

The graphical presentation of a transformation $\alpha \in \mathcal{PT}_n$ via Γ_α is very transparent, but also rather space consuming. For a plain mathematical text, it would be very useful to have some space-saving alternative. For permutations this is known as the *cyclic* notation and can be easily described by the following example:

Example 1.3.1 For the permutation

$$
\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 9 & 8 & 15 & 2 & 10 & 1 & 14 & 4 & 7 & 5 & 6 & 11 & 13 & 3 & 12 \end{pmatrix}
$$

the graph Γ_α has the following form:

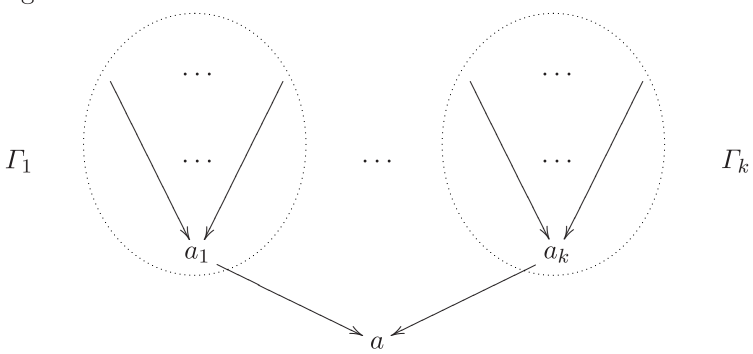


Using the notation for cycles, introduced on page 4 in the paragraph after Exercise 1.2.2, we may write

$$\alpha = (1, 9, 7, 14, 3, 15, 12, 11, 6)(2, 8, 4)(5, 10)(13).$$

Clearly the above notation is not uniquely defined. Writing a cycle we can start from each of its vertices. Moreover, the order of cycles in the cyclic notation can also be chosen in an arbitrary way. In this subsection, we would like to generalize this notation to be able to use it for all elements of $\mathcal{PT}_n$. A very good hint how to do this is given by Theorem 1.2.9, which roughly says that we only have to find a nice notation for trees with sinks. We call our notation *linear* and will define it recursively.

Assume for the moment that the graph Γ_α, where $\alpha \in \mathcal{PT}_n$, is a tree with the sink a. If a is the only vertex of Γ, we shall write $\Gamma_\alpha = [a]$ (or simply $\alpha = [a]$). If Γ_α contains some other vertices, then it has to have the following form:



(1.5)

For $i = 1, \ldots, k$, the subgraph Γ_i of the graph (1.5) above is a tree with the sink a_i and has strictly less vertices than Γ_α. Assume that we already have the linear notation $\tilde{\Gamma}_i$ for Γ_i, $i = 1, \ldots, k$. Then the *linear notation* for Γ_α (and α) is defined recursively as follows

$$\Gamma_\alpha = [\tilde{\Gamma}_1, \tilde{\Gamma}_2, \ldots, \tilde{\Gamma}_k; a].$$

This defines the notation for the elements given by Theorem 1.2.9(i).

Assume now that Γ_α is connected and given by Theorem 1.2.9(ii). Then Γ_α is the union of some cycle, say $(a_1, \ldots, a_k)$, with certain disjoint trees Γ_i with sinks a_i, $i = 1, \ldots, k$. Let $\tilde{\Gamma}_i$, $i = 1, \ldots, k$, be the corresponding linear notation. In this case, we define the linear notation for Γ_α (and α) as follows

$$\Gamma_\alpha = (\tilde{\Gamma}_1, \tilde{\Gamma}_2, \ldots, \tilde{\Gamma}_k).$$

Finally, for any $\alpha \in \mathcal{PT}_n$, we define the linear notation for Γ_α (and α) to be the product of linear notation over all connected components of Γ_α, written in an arbitrary order. In the same way as the classical cycle notation for permutations, the linear notation for (partial) transformartions is unique only up to permutation of certain components of the notation. Namely, the connected components can be written in an arbitrary order, and on each step of the recursive procedure the order of the components $\tilde{\Gamma}_1, \dots, \tilde{\Gamma}_k$ can also be chosen arbitrarily.

Note that, by the above definition, the ordinary cycle $(a_1, a_2, \dots, a_k)$ is denoted by $([a_1], [a_2], \dots, [a_k])$. This is of course not very practical, so to avoid this unnecessary complication inside the notation for cycles (but not for trees with sinks!) we shall usually skip the brackets "[ ]" surrounding *trivial* trees with sinks. Sometimes, if n is fixed, one can also skip all loops (i.e., the elements of the form $(x) = ([x])$). This just means that all $x \in \mathbf{N}$, which do not appear in the notation, correspond to loops. It is clear that this does not give rise to any confusion, moreover, it restores the original notation for usual cycles and permutations.

Example 1.3.2 For the transformation α from Example 1.2.1 we have

$$\alpha = ([[[[15];14],[2];9],[[[10];11],[6];16];13],12,4)(1,8)[[3],[5];7].$$

1.4 Addenda and Comments

1.4.1 To use graphs for presentation of transformations was proposed by Suschkewitsch in [Su1].

1.4.2 Let $\alpha \in \mathcal{PT}_n$. The element $x \in \mathbf{N}$ satisfying $\alpha(x) = x$ is usually called a *fixed point* of α. If α is a permutation, then all the connected components of Γ_α are cycles. Fixed points of α correspond to cycles of length 1. In the cyclic notation for α such cycles are usually omitted (for the identity element one thus has to use a special notation, for example ε). If after such simplification the cyclic notation for α contains only one cycle, say of length k, the α is usually called a *cycle of length k*. Cycles of length 2 are called *transpositions*.

1.4.3 An alternative "linear" notation for total transformations was proposed in [AAH]. Although [AAH] works only with total transformations it is fairly straightforward to generalize their notation to cover all partial transformations. Roughly speaking the [AAH]-notation reduces to listing the trajectories of all vertices of the graph Γ_α. If we know the trajectory $x_0 = x, x_1, x_2, \dots$ of the vertex x, then of course we know the trajectories of each of the vertices x_1, $x_2, \dots$, so we can omit the latter ones. A trajectory is denoted by $[a_1, a_2, \dots, a_k | a_j]$, where $1 \leq j \leq k$. This means the following:

- If $j < k$, the trajectory $a_1, a_2, \ldots$ of the vertex a_1 terminates at the cycle $(a_j, a_{j+1}, \ldots, a_k)$
- If $j = k$ and a_k does not occur previously, then the trajectory $a_1, a_2, \ldots$ of the vertex a_1 terminates at the cycle (a_k)
- If $j = k$ and a_k does occur previously, it means that we already know the trajectory of a_k, in this case the trajectory of the vertex a_1 is obtained by attaching $a_1, a_2, \ldots, a_k$ to the known trajectory of a_k

On each step we choose any vertex, say a, whose trajectory is not yet written down, and we write down the trajectory of a until we either reach a vertex, whose trajectory is already written down, or we terminate the trajectory of a in some cycle. To make the notation as short as possible, on each step one should try to choose a new vertex, which is not a head of any arrow in Γ_α.

For example, for the transformation

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 8 & 9 & 7 & 13 & 7 & 16 & 7 & 1 & 13 & 11 & 16 & 4 & 12 & 9 & 14 & 13 \end{pmatrix}$$

the [AAH]-notation for α will have the following form:

$$[15, 14, 9, 13, 12, 4|13][10, 11, 16, 13|13][6, 16|16][2, 9|9][8, 1|8][3, 7|7][5, 7|7].$$

For comparison, our notation for α looks as follows:

$$([[[[15]; 14], [2]; 9], [[[10]; 11], [6]; 16]; 13], 12, 4)(1, 8)([[3], [5]; 7]).$$

From our point of view the [AAH]-notation has some disadvantages, namely,

- The cyclic notation for permutations is not a partial case of the [AAH]-notation
- The [AAH]-notation is long, that is, it always contains elements occurring more than one time
- The [AAH]-notation is by far not unique, even up to permutations of certain components of this notation

Another disadvantage of the [AAH]-notation, related to the composition of transformations, will be discussed in 2.9.3. An advantage of the [AAH]-notation in comparison to our notation is that it contains less brackets.

1.4.4 In [Ka] it was shown that the number of those $\alpha \in \mathcal{T}_n$ for which Γ_α is connected equals

$$n! \sum_{k=0}^{n-1} \frac{n^{k-1}}{k!}$$

This can be proved, for example, in the following way:

From Theorem 1.2.9 it follows that Γ_α is connected if and only if it is a union of a forest of trees with sinks with a cycle on the set of all sinks. For each such α and for each $i \geq 0$ we define the set $N_\alpha^{(i)}$ as the set of all $x \in \mathbf{N}$ such that the first element from the cycle occurs in $\mathrm{tr}_\alpha(x)$ on step i. Obviously, $N_\alpha^{(0)}$ consists of the elements of our cycle and $\mathbf{N} = \cup_{i\geq 0} N_\alpha^{(i)}$ is a disjoint union. Furthermore, $\alpha(N_\alpha^{(i)}) \subset N_\alpha^{(i-1)}$ for $i > 0$. The number of those α, for which $|N_\alpha^{(i)}| = n_i$, $0 \leq i \leq t$, and such that $\sum_{i=0}^{t} n_i = n$ equals

$$\binom{n}{n_0, n_1, \dots, n_t}(n_0-1)!\, n_0^{n_1} n_1^{n_2} \dots n_{t-1}^{n_t} \tag{1.6}$$

(here the polynomial coefficient $\binom{n}{n_0,n_1,\dots,n_t}$ gives the number of ordered partitions of $\mathbf{N}$ into blocks with cardinalities $n_0, \dots, n_t$, respectively; $(n_0-1)!$ is the number of ways to form a cycle out of n_0 elements; and $n_i^{n_{i+1}}$ is the number of maps from the block with n_{i+1} elements to the block with n_i elements). We can rewrite (1.6) as follows

$$\frac{n!}{n_0} \cdot \frac{n_0^{n_1}}{n_1!} \cdot \frac{n_1^{n_2}}{n_2!} \dots \frac{n_{t-1}^{n_t}}{n_t!}. \tag{1.7}$$

Now to find the number X_{n_0} of those $\alpha \in \mathcal{T}_n$ for which Γ_α is connected and contains a cycle of length n_0 one has to add up all summands of the form (1.7) for all $t \leq n - n_0$ and all decompositions $n_1 + \dots + n_t = n - n_0$. Let $n - n_0 = k$. Then

$$\frac{n^{k-1}}{k!} = \frac{(n_0+k)^{k-1}}{k!} = \sum_{i=0}^{k-1} \frac{1}{k!} \cdot \frac{(k-1)!}{i!(k-1-i)!} \cdot n_0^i \cdot k^{k-1-i} = \\ = \sum_{n_1=1}^{k} \frac{n_0^{n_1-1}}{(n_1-1)!} \cdot \frac{k^{k-n_1}}{(k-n_1)!},$$

where for the last equality we substituted i by $n_1 - 1$. Continuing in the same way we get

$$\frac{n^{k-1}}{k!} = \\ = \sum_{n_1=1}^{k} \frac{n_0^{n_1-1}}{(n_1-1)!} \cdot \sum_{n_2=1}^{k-n_1} \frac{n_1^{n_2-1}}{(n_2-1)!} \cdots \sum_{n_t=1}^{k-n_1-\dots-n_{t-1}} \frac{n_{t-1}^{n_t-1}}{(n_t-1)!} \cdot n_t^{-1} = \\ = \frac{1}{n_0} \sum_{n_1=1}^{k} \frac{n_0^{n_1}}{n_1!} \sum_{n_2=1}^{k-n_1} \frac{n_1^{n_2}}{n_2!} \cdots \sum_{n_t=1}^{k-n_1-\dots-n_{t-1}} \frac{n_{t-1}^{n_t}}{n_t!} = \\ = \sum_{n_1+\dots+n_t=n-n_0} \frac{1}{n_0} \cdot \frac{n_0^{n_1}}{n_1!} \cdot \frac{n_1^{n_2}}{n_2!} \dots \frac{n_{t-1}^{n_t}}{n_t!}. \tag{1.8}$$

Hence $X_{n_0} = n! \cdot \frac{n^{k-1}}{k!}$, where $k = n - n_0$. Since $1 \leq n_0 \leq n$, the final answer is now computed as follows

$$\sum_{n_0=1}^{n} X_{n_0} = n! \sum_{k=0}^{n-1} \frac{n^{k-1}}{k!}.$$

1.5 Additional Exercises

1.5.1 Let $\mathbb{N}$ denote the set of all positive integers. Give an example of a transformation $\alpha : \mathbb{N} \to \mathbb{N}$ such that

(a) α is injective but not surjective.

(b) α is surjective but not injective.

1.5.2 Prove that $\lim_{n\to\infty} \frac{|\mathcal{PT}_n|}{|\mathcal{T}_n|} = e$.

1.5.3 Directed graphs $\Gamma_i = (V_i, E_i)$, $i = 1, 2$, are called *isomorphic* provided that there exists a bijection $\varphi : V_1 \to V_2$ which induces a bijection from E_1 to E_2. Compute the number of pairwise nonisomorphic graphs Γ_α, where

(a) $\alpha \in \mathcal{T}_3$.

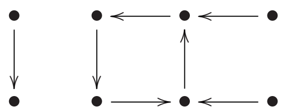
(b) $\alpha \in \mathcal{T}_4$.

(c) $\alpha \in \mathcal{PT}_2$.

(d) $\alpha \in \mathcal{PT}_3$.

(e) $\alpha \in \mathcal{PT}_4$.

1.5.4 Find the number of those partial transformations $\alpha \in \mathcal{PT}_8$, whose graphs are isomorphic to the following graph:



1.5.5 For $\alpha \in \mathcal{PT}_n$ characterize $\mathrm{dom}(\alpha)$, $\mathrm{im}(\alpha)$, $\overline{\mathrm{dom}}(\alpha)$, $\mathrm{rank}(\alpha)$, and $\mathrm{def}(\alpha)$ in terms of Γ_α.

1.5.6 Compute the number of those $\alpha \in \mathcal{T}_n$ (resp. $\alpha \in \mathcal{PT}_n$) for which $\mathrm{im}(\alpha)$

(a) Does not contain given elements a_1, a_2,..., a_k

(b) Contains given elements a_1, a_2,..., a_k

(c) Coincides with the given set $\{a_1, a_2, \dots, a_k\}$

1.5.7 Prove that the number of those $\alpha \in \mathcal{PT}_n$, for which Γ_α is a tree with a sink, equals n^{n-1}.

1.5.8 Prove that the number of those $\alpha \in \mathcal{PT}_n$, for which Γ_α does not contain cycles, equals $\sum_{k=1}^{n} \binom{n-1}{k-1} n^{n-k}$.

1.5.9 (a) Find the number of those $\alpha \in \mathcal{T}_n$ which fix at least one (resp. exactly one) element (that is, $\alpha(x) = x$ for at least one or exactly one element $x \in \mathbf{N}$, respectively).

(b) The same problem for $\mathcal{PT}_n$.

1.5.10 Let $\Gamma = (V, E)$ be a directed graph. Consider the set $\mathcal{X}$, which consists of all possible unordered partitions of V into disjoint unions of nonempty subsets V_is such that for each $i \neq j$ the graph Γ does not contain any arrow from V_i to V_j. The set $\mathcal{X}$ is partially ordered in the natural way with respect to inclusions of components of partitions. Prove that the partition of V, which corresponds to the partition of Γ into connected components, is the minimum of $\mathcal{X}$.

1.5.11 Let $\Gamma = (V, E)$ be a directed graph. Consider the set $\mathcal{Y}$, which consists of all possible unordered partitions of V into disjoint unions of nonempty subsets V_is such that for each i we have that the subgraph $(V_i, (V_i \times V_i) \cap E)$ is connected. The set $\mathcal{Y}$ is partially ordered in the natural way with respect to inclusions of components of partitions. Prove that the partition of V, which corresponds to the partition of Γ into connected components, is the maximum of $\mathcal{Y}$.

1.5.12 For $\alpha \in \mathcal{T}_n$, let $\mathfrak{t}_k(\alpha)$ denote the number of those $x \in \mathbf{N}$ for which $|\{y \in \mathbf{N} : \alpha(y) = x\}| = k$. Prove that

(a) $\displaystyle\sum_{k=0}^{n} \mathfrak{t}_k(\alpha) = n,$

(b) $\displaystyle\sum_{k=0}^{n} k\mathfrak{t}_k(\alpha) = n.$

1.5.13 For $\alpha \in \mathcal{PT}_n$ let $\mathfrak{t}_k(\alpha)$ denote the number of those $x \in \mathbf{N}$ for which $|\{y \in \mathbf{N} : \alpha(y) = x\}| = k$. Prove that

(a) $\displaystyle\sum_{k=0}^{n} \mathfrak{t}_k(\alpha) = n,$

(b) $\displaystyle\sum_{k=0}^{n} k\mathfrak{t}_k(\alpha) \leq n.$

Chapter 2

The Semigroups $\mathcal{T}_n$, $\mathcal{PT}_n$, and $\mathcal{IS}_n$

2.1 Composition of Transformations

Let X and Y be two sets. A *mapping* from X to Y is an array of the form

$$f = \begin{pmatrix} x \\ f(x) \end{pmatrix}_{x \in X},$$

where all $f(x) \in Y$. This is usually denoted by $f : X \to Y$. The element $f(x)$ is called the *value* of the mapping f at the element x. A transformation, as defined in Sect. 1.1, is just a mapping from a set to itself. Let now X, Y, Y', Z be sets such that $Y \subset Y'$ and let $f : X \to Y$ and $g : Y' \to Z$ be two mappings. In this situation, we can define the *product* or the *composition* gf of f and g by the following rule: The composition gf is the mapping from X to Z such that for all $x \in X$ we have $(gf)(x) = g(f(x))$. In particular, we can always compose two total transformations of the same set and the result will be a total transformation of this set.

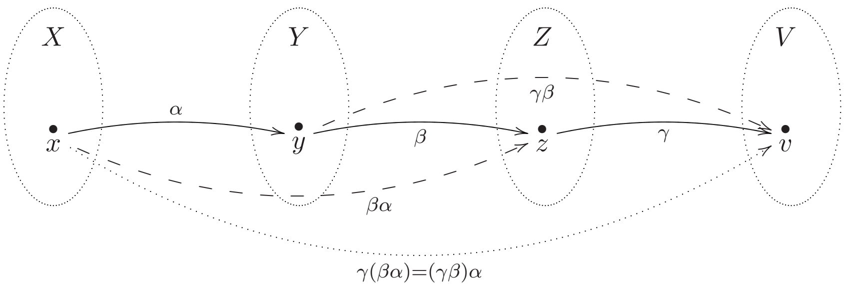
The above definition admits a straightforward generalization to partial mappings. A *partial mapping* from X to Y is a mapping $\alpha : X' \to Y$, where $X' \subset X$. In this case, we say that the partial mapping α is *defined* on elements from X'. Again, a partial transformation, as defined in Sect. 1.1, is a partial mapping from a set to itself. One usually abuses notation and writes $\alpha : X \to Y$ just emphasizing that α is a partial mapping. Let $\alpha : X \to Y$ and $\beta : Y \to Z$ be two partial mappings. We define their *product* or *composition* $\beta\alpha$ as the partial mapping, defined on all those $x \in X$ for which α and β are defined on the elements x and $\alpha(x)$, respectively; on such x the value of $\beta\alpha$ is given by $(\beta\alpha)(x) = \beta(\alpha(x))$. In particular, we can always compose two partial transformations of the same set and the result will be another

O. Ganyushkin, V. Mazorchuk, *Classical Finite Transformation Semigroups,* Algebra and Applications 9, DOI: 10.1007/978-1-84800-281-4_2,

partial transformation of this set. We also note that the definition of the composition of total transformations is just a special case of that of partial transformations.

Proposition 2.1.1 *The composition of (partial) mappings is associative, that is, if α, β, and γ are partial mappings, then the composition $\gamma(\beta\alpha)$ is defined if and only if the composition $(\gamma\beta)\alpha$ is defined, and if they both are defined, we have $\gamma(\beta\alpha) = (\gamma\beta)\alpha$.*

Proof. Follows immediately from the following picture:



□

Associativity of the composition of partial transformations naturally leads to the notion of a semigroup. Let S be a nonempty set, and let $\cdot : S \times S \to S$ be a *binary operation* on S. Then $(S, \cdot)$ is called a *semigroup* provided that $\cdot$ is *associative*, that is, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in S$. To simplify the notation, in the case when the operation $\cdot$ is clear from the context one usually writes S for $(S, \cdot)$. Furthermore, one usually writes ab instead of $a \cdot b$.

Exercise 2.1.2 Let $(S, \cdot)$ be a semigroup. Show that the value of the product $a_1 a_2 \cdots a_n$, where all $a_i \in S$, does not depend on the way of computing it (that is, of putting brackets into this product).

Let $(S, \cdot)$ be a semigroup. From Exercise 2.1.2 it follows that for every $a \in S$ we have a well-defined element $a^k = \underbrace{a \cdot a \cdots a}_{k\text{times}}$. The number of elements in S is called the *cardinality* of S and is denoted by $|S|$.

By Proposition 2.1.1, in both $\mathcal{T}_n$ and $\mathcal{PT}_n$ the composition of (partial) transformations is an associative operation. Hence we have:

Proposition 2.1.3 *Both $\mathcal{T}_n$ and $\mathcal{PT}_n$ are semigroups with respect to the composition of (partial) transformations.*

The semigroup $\mathcal{T}_n$ is called the *full transformation semigroup* on the set $\mathbf{N}$ or the *symmetric semigroup* of all transformations of $\mathbf{N}$. The semigroup $\mathcal{PT}_n$ is called the *semigroup of all partial transformations* on $\mathbf{N}$.

A nonempty subset T of a semigroup $(S, \cdot)$ is called a *subsemigroup* of S provided that T is closed with respect to $\cdot$ (that is, $a \cdot b \in T$ as soon as $a, b \in T$). Obviously, in this case, T itself is a semigroup with respect to the restriction of the operation $\cdot$ to T. The fact that T is a subsemigroup of S is usually denoted by $T < S$.

Exercise 2.1.4 Show that for arbitrary $\alpha, \beta \in \mathcal{PT}_n$ the following is true:

(a) $\mathrm{dom}(\beta\alpha) \subset \mathrm{dom}(\alpha)$

(b) $\mathrm{im}(\beta\alpha) \subset \mathrm{im}(\beta)$

(c) $\mathrm{rank}(\beta\alpha) \leq \min(\mathrm{rank}(\alpha), \mathrm{rank}(\beta))$

2.2 Identity Elements

An element e of a semigroup S is called a *left* or a *right identity* provided that $ea = a$, or $ae = a$, respectively, for all $a \in S$. An element e, which is a left and a right identity at the same time, is called a *two-sided identity* or simply an *identity.*

If S contains some left identity e_{l} and some right identity e_{r} we have $e_{\mathrm{l}} = e_{\mathrm{l}} \cdot e_{\mathrm{r}} = e_{\mathrm{r}}$ and hence these two elements coincide. Hence in this case S contains a unique identity element, which is, moreover, a two-sided identity. However, a semigroup may contain many different left identities or many different right identities (see Exercise 2.10.2). It is possible for a semigroup to contain neither left nor right identities. An example of such a semigroup is the semigroup $(\mathbb{N}, +)$. Another example is the semigroup $\{2, 3, 4, \dots\}$ with respect to the ordinary multiplication.

A semigroup which contains a two-sided identity element is called a *monoid.* The absence of an identity element can be easily repaired in the following way.

Proposition 2.2.1 *Each semigroup can be extended to a monoid by adding at most one element.*

Proof. Let $(S, \cdot)$ be a semigroup. If S contains an identity, we have nothing to prove. If S does not contain any identity element, consider the set $S^1 = S \cup \{1\}$, where $1 \notin S$. Define the binary operation $*$ on S^1 as follows: For $a, b \in S^1$ set

$$a * b = \begin{cases} a \cdot b, & a, b \in S; \\ a, & b = 1; \\ b, & a = 1. \end{cases}$$

A direct calculation shows that $*$ is associative, hence S^1 is a semigroup. Furthermore, from the definition of $*$ we have that 1 is the identity element in S^1. Moreover, the restriction of the operation $*$ to S coincides with the original operation $\cdot$. Hence S is a subsemigroup of S^1. $\square$

Denote by $\varepsilon_n : \mathbf{N} \to \mathbf{N}$ the *identity* transformation

$$\varepsilon_n = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}.$$

If n is clear from the context we shall sometimes write ε instead of ε_n. The following statement is obvious.

Proposition 2.2.2 *The transformation ε_n is the (two-sided) identity element in both $\mathcal{T}_n$ and $\mathcal{PT}_n$. In particular, both, $\mathcal{T}_n$ and $\mathcal{PT}_n$, are monoids.*

Let S be a monoid with the identity element 1. An element $a \in S$ is called *invertible* or a *unit* provided that there exists $b \in S$ such that $ab = ba = 1$. Such an element b, if it exists, is unique. Indeed, assume that b_1 and b_2 are different such elements, then

$$b_1 = b_1 \cdot 1 = b_1(ab_2) = (b_1a)b_2 = 1 \cdot b_2 = b_2.$$

The element b is called the *inverse* of a and is denoted by a^{-1}. Note that if b is the inverse of a, then a is the inverse of b. In other words, $(a^{-1})^{-1} = a$. The set of all invertible elements of the monoid S is denoted by S^*. Note that $1 \in S^*$ since $1 \cdot 1 = 1$. In particular, S^* is not empty.

The above terminology and notation deserve some explanation. Usually the operation in an abstract semigroup is thought of as a multiplication. Since the element 1 is the identity element in such multiplicative semigroups as $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$, it is natural to denote the identity element of an abstract semigroup by the same symbol 1. This also justifies the notions "unit" and "inverse." However, there are many semigroups where the operation is not the multiplication, for example the semigroup $(\mathbb{Z}, +)$. The identity element in this semigroup is the number 0 and not the number 1. And the inverse of the number $n \in \mathbb{Z}$ is the number $-n$ and not the number n^{-1} (note that the latter one is not always defined, and when it is defined, it is not an integer in general).

A monoid in which each element has an inverse is called a *group*.

Proposition 2.2.3 *Let S be a monoid with the identity element* 1*. Then S^* is a group.*

Proof. Obviously, if $a \in S^*$, then $a^{-1} \in S^*$ as well. If $a, b \in S^*$, then we have

$$ab \cdot b^{-1}a^{-1} = a \cdot bb^{-1} \cdot a^{-1} = a \cdot 1 \cdot a^{-1} = aa^{-1} = 1.$$

Analogously one shows that $b^{-1}a^{-1} \cdot ab = 1$ and hence $b^{-1}a^{-1} = (ab)^{-1}$. In particular, $ab \in S^*$. Thus S^* is a submonoid of S and each element of S^* has an inverse in S^*. The claim follows. □

Proposition 2.2.4 *Let $\alpha \in \mathcal{T}_n$, or $\alpha \in \mathcal{PT}_n$. Then α is invertible if and only if α is a permutation on $\mathbf{N}$.*

Proof. Assume that α is invertible and β is a (partial) transformation such that $\alpha\beta = \beta\alpha = \varepsilon$. Note that $\mathrm{dom}(\varepsilon) = \mathbf{N}$. Hence Exercise 2.1.4(a) implies $\mathrm{dom}(\alpha) = \mathbf{N}$. Further, if $x, y \in \mathbf{N}$ are such that $x \neq y$, then $\varepsilon(x) \neq \varepsilon(y)$. If $\alpha(x) = \alpha(y)$, we would get $\varepsilon(x) = \beta(\alpha(x)) = \beta(\alpha(y)) = \varepsilon(y)$, a contradiction. This means that $\alpha(x) \neq \alpha(y)$. Hence α is everywhere defined and injective and thus is a permutation by Proposition 1.1.3.

Conversely, if

$$\alpha = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$$

is a permutation, the element

$$\alpha = \begin{pmatrix} i_1 & i_2 & \cdots & i_n \\ 1 & 2 & \cdots & n \end{pmatrix}$$

is a permutation as well and a direct computation shows that $\alpha\beta = \beta\alpha = \varepsilon$, that is, α is invertible. □

The group $\mathcal{T}_n^* = \mathcal{PT}_n^*$ of all permutations on $\mathbf{N}$ is called the *symmetric group* on $\mathbf{N}$ and is denoted by $\mathcal{S}_n$.

2.3 Zero Elements

An element 0 of a semigroup S is called a *left* or a *right zero* provided that $0a = 0$, or $a0 = 0$, respectively, for all $a \in S$. An element 0 which at the same time is a left and a right zero, is called a *two-sided zero* or simply a *zero*.

If S contains some left zero 0_l and some right zero 0_{r}, we have $0_{\mathrm{l}} = 0_{\mathrm{l}} \cdot 0_{\mathrm{r}} = 0_{\mathrm{r}}$ and hence these two elements coincide. Hence in this case S contains a unique zero element, which is, moreover, a two-sided zero. The analog of Proposition 2.2.1 is the following statement.

Proposition 2.3.1 *Each semigroup can be extended to a semigroup with zero by adding at most one element.*

Proof. Let $(S, \cdot)$ be a semigroup. If S contains a zero, we have nothing to prove. Otherwise, consider the set $S^0 = S \cup \{0\}$, where $0 \notin S$. Define the binary operation $*$ on S^0 as follows: For $a, b \in S^0$ set

$$a * b = \begin{cases} a \cdot b, & a, b \in S; \\ 0, & b = 0; \\ 0, & a = 0. \end{cases}$$

A direct calculation shows that $*$ is associative, hence S^0 is a semigroup. Furthermore, from the definition of $*$ we have that 0 is the zero element in S^0 and that the restriction of $*$ to S coincides with $\cdot$. Hence S is a subsemigroup of S^0. □