# Probability Measures on Semigroups

## Convolution Products, Random Walks, and Random Matrices

## THE UNIVERSITY SERIES IN MATHEMATICS



Recent volumes in the series:

**THE CLASSIFICATION OF FINITE SIMPLE GROUPS**
Daniel Gorenstein
VOLUME 1: GROUPS OF NONCHARACTERISTIC 2 TYPE

**COMPLEX ANALYSIS AND GEOMETRY**
Edited by Vincenzo Ancona and Alessandro Silva

**ELLIPTIC DIFFERENTIAL EQUATIONS AND OBSTACLE PROBLEMS**
Giovanni Maria Troianiello

**FINITE SIMPLE GROUPS: An Introduction to Their Classification**
Daniel Gorenstein

**AN INTRODUCTION TO ALGEBRAIC NUMBER THEORY**
Takashi Ono

**MATRIX THEORY: A Second Course**
James M. Ortega

**PROBABILITY MEASURES ON SEMIGROUPS: Convolution Products, Random Walks, and Random Matrices**
Göran Högnäs and Arunava Mukherjea

**A SCRAPBOOK OF COMPLEX CURVE THEORY**
C. Herbert Clemens

**TOPICS IN NUMBER THEORY**
J. S. Chahal

**VARIATIONS ON A THEME OF EULER**
**Quadratic Forms, Elliptic Curves, and Hopf Maps**
Takashi Ono

# Probability Measures on Semigroups

## Convolution Products, Random Walks, and Random Matrices

**Göran Högnäs**
*Åbo Akademi University*
*Åbo, Finland*

**Arunava Mukherjea**
*University of South Florida*
*Tampa, Florida*

Springer Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

Högnäs, Göran.
Probability measures on semigroups : convolution products, random walks, and random matrices / Göran Högnäs and Arunava Mukherjea.
p. cm. -- (University series in mathematics)
Includes bibliographical references (p. - ) and index.
ISBN 978-1-4757-2390-8 ISBN 978-1-4757-2388-5 (eBook)
DOI 10.1007/978-1-4757-2388-5
1. Probability measures. 2. Semigroups. I. Mukherjea, Arunava, 1941- . II. Title. III. Series: University series in mathematics (Plenum Press)
QA273.6.H665 1995
519.2'6--dc20 95-9418
CIP

ISBN 978-1-4757-2390-8


Originally published by Plenum Press, New York in 1995
Softcover reprint of the hardcover 1st edition 1995

10 9 8 7 6 5 4 3 2 1

# Preface

A *Scientific American* article on chaos, see Crutchfield *et al.* (1986), illustrates a very persuasive example of recurrence. A painting of Henri Poincaré, or rather a digitized version of it, is stretched and cut to produce a mildly distorted image of Poincaré. The same procedure is applied to the distorted image and the process is repeated over and over again on the successively more and more blurred images. After a dozen repetitions nothing seems to be left of the original portrait. Miraculously, structured images appear briefly as we continue to apply the distortion procedure to successive images. After 241 iterations the original picture reappears, unchanged!

Apparently the pixels of the Poincaré portrait were moving about in accordance with a strictly deterministic rule. More importantly, the set of all pixels, the whole portrait, was *transformed* by the distortion mechanism. In this example the transformation seems to have been a *reversible* one since the original was faithfully recreated.

It is not very farfetched to introduce a certain amount of *randomness* and *irreversibility* in the above example. Think of a random miscoloring of some pixels or of inadvertently giving a pixel the color of its neighbor.

The methods in this book are geared towards being applicable to the asymptotics of such transformation processes. The transformations form a semigroup in a natural way; we want to investigate the long-term behavior of random elements of this semigroup.

To be more specific, let us consider a sequence of independent and identically distributed random variables $X_0, X_1, X_2, \ldots$ taking values in a set of affine maps from $R^d$ into $R^d$, that is, maps of the form $f(x) = Ax + B$, where $B$ and $x$ are $d \times 1$ column vectors and $A$ is a $d \times d$ real matrix. Since $f$ can be identified with the $(d+1) \times (d+1)$ matrix $\begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}$, the random variables

$X_i$s can also be regarded as $(d+1) \times (d+1)$ random matrices; thus, $\mu$, the distribution of $X_i$, is a probability measure on the set of $(d+1) \times (d+1)$ matrices of the form $\begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}$. Let $S$ be the closed (with usual topology) multiplicative semigroup generated by the support of $\mu$. Then the study of the random walks $Y_n$, $Y_n = X_0 X_1 \dots X_n$ with values in $S$ and distribution $\mu^n$ (the $n$th convolution power of $\mu$), and the set of recurrent states of $(Y_n)$ become relevant in the context of the so-called iterated function systems introduced by Barnsley and his colleagues [see Barnsley (1988)].

Let us briefly discuss another example.

Suppose that we are monitoring a random system with two states denoted 0 and 1. Let

$$0\,1\,0\,0\,0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,0\,1\,0\,0\,1\,1\,0\,0\,1\,1\,1\,1\,0\,0\,0$$

and

$$1\,0\,1\,1\,1\,0\,0\,1\,0\,1\,1\,0\,0\,1\,1\,1\,0\,1\,1\,1\,1\,0\,0\,1\,1\,1\,1\,0\,0\,0$$

be observed time series of the successive states of the system. The observations seem rather like a record of independent coin tosses, with 0 for heads and 1 for tails, say. Viewed as a Markov chain on the two-state state space $X = \{0,1\}$ our process would have the transition probability matrix

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Let us assume, however, that the above time series are *concurrent.* Then another interpretation imposes itself: the state space is subjected to a succession of random transformations. (The first two transformations are transpositions, 0 and 1 just trade places. At the third and fourth steps the identity map is at work. A sequence of transpositions and identities then follows, but at step 19 everything is mapped onto the state 1. From then on the two paths are identical.) The transformations are the four possible mappings of $X$ into itself, the identity $\iota$, the transposition $\tau$, and the two constant mappings 0 and 1. The transition matrix $P$ is then a convex combination of matrices representing those transformations:

$$P = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + c\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

where $a$, $b$, $c$, $d$ are nonnegative numbers with $a+b+c+d=1$.

Thus, a natural way to analyze our observed time series is to think of them as emanating from an independent, identically distributed sequence of mappings

of the state space into itself, or, in other words, a *random walk on the transformations* of $X$.

A Markov chain on a finite state space can always be regarded in this way. Its transition matrix $P$ is a convex combination of 0-1 matrices representing mappings of the state space $X$ into itself. (If $P$ is doubly stochastic we can write it uniquely as a convex combination of permutation matrices, this is the celebrated Birkhoff theorem.) The corresponding result is true even for a large class of Markov chains on a topological space $X$, see Kifer (1986), Chapter 1.

To consider an example in the context of particle systems, let $V$ be an arbitrary countably infinite set (with discrete topology), and let $\Gamma$ denote the semigroup of functions $f : V \to V$ under composition. We can then identify each $f$ in $\Gamma$ with an infinite 0-1 stochastic matrix $A_f$ such that

$$(A_f)_{ij} = 1 \quad \text{if and only if} \quad f(i) = j.$$

By a configuration $\eta$ of $V$, we mean a nonnegative integer valued function on $V$ such that

$$\sum_{x \in V} \eta(x) < \infty.$$

The idea is that $\eta(j)$ is the number of particles that occupy the site $j \in V$, and that when we apply the mapping $f : V \to V$, all these particles move to the site $f(j)$, and the configuration changes to $\eta \cdot A_f$, where

$$(\eta \cdot A_f)(x) = \sum_y \eta(y) \cdot \delta_{f(y)}(x).$$

The new configuration has at site $x$ all the particles that the map $f$ has sent to the site $x$ from the sites of the original configuration. Thus, to study the random motions of finite systems of particles on $V$, without births or deaths, where each site may be occupied by a finite number of particles, and all particles at a particular site move together, one needs to study the random transformations $F$ (that is, the infinite random stochastic matrices $A_F$). Instead of studying the different configurations, we study a sequence of independent identically distributed countably infinite stochastic matrices, and among other things, will be interested in gaining some insights in the limiting laws of products of these matrices.

To mention yet another context where probability measures on countable semigroups have been found useful, we mention the paper of Hansel and Perrin (1983), where the authors utilized the structure of an idempotent probability measure on a semigroup in order to have some insights in certain problems in coding theory.

It is also relevant to mention that Ruzsa (1994) utilized his results on weak* convergence of the sequence $\mu_1 * \mu_2 * \ldots * \mu_n$, where the $\mu_i$s are probability measures on a countable semigroup, in proving a generalization of a result in

number theory due to Davenport and Erdös (1936). This last mentioned result simply says that every multiplicative ideal $A$ of the set $N$ of positive integers has a logarithmic density, that is,

$$\lambda(A) = \lim_{n\to\infty} \frac{1}{\log n} \sum_{\substack{a\in A \\ a\le n}} \frac{1}{a},$$

exists. Note that for a set $A \subset N$, its logarithmic density $\lambda$ may exist while its asymptotic density $d$, given by

$$d(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{\substack{a\in A \\ a\le n}} 1,$$

may not exist. [It is well known, however, that $\lambda(A)$ exists whenever $d(A)$ does, and then $\lambda(A) = d(A)$.] Ruzsa's result says the following: if $f$ is a homomorphism from the multiplicative semigroup of integers to a commutative semigroup $H$, then for every $h \in H$, the set $\{n \in N : f(n) = h\}$ has a logarithmic density.

Let us finally mention, before we go to the text proper, that abstract semigroup theory was of crucial importance in developing the methods used in Högnäs and Mukherjea (1980) to study the set of recurrent states of a random walk taking values in $n \times n$ real matrices.

Now to describe the contents of this book, let us say that here we make an attempt to present up-to-date information in the theory of weak convergence of convolution products of probability measures on semigroups (Chapter 2), the theory of random walks with values in semigroups (Chapter 3), and applications of the preceding theories to products of random matrices (Chapter 4). Chapter 1 contains essentials of abstract semigroup theory along with its application to concrete semigroups of matrices. Chapter 1, while it contains many important results from abstract semigroup theory, is not designed to cover semigroup theory in depth, and as such contains mostly those results and concepts which are needed for an understanding of later chapters.

To restrict the size of the book, we have often been biased towards presenting only those results which are new, surprising, useful, and interesting in the context of semigroups. However, certain results have been presented in groups rather than semigroups, and this has been done whenever the corresponding semigroup situation is not very clear or so far only partially solved. Thus, we have discussed concentration functions (in Chapter 2) only in the context of groups.

A graduate student familiar with material covered in standard courses (in a typical American university) in probability theory, measure theory, group theory, topology, and linear algebra should not have any difficulty in following this book

on his/her own. A two-semester special topics course on weak convergence and random walks can be based on the material covered in this book.

For ease of reading, let us also mention that the theorems, lemmas, propositions, and corollaries in each chapter of this book are numbered consecutively in the order in which they appear. Thus, Proposition 2.18 follows Corollary 2.17, Lemma 2.20 follows Proposition 2.19, and Theorem 2.23 follows Lemma 2.22.

Finally, let us express our gratitude to many of our colleagues and friends for assisting us in various ways (in the writing of this book) through discussions and actual collaboration in research. We are specially indebted to Herbert Heyer, Karl Hofmann, Imre Ruzsa, Murray Rosenblatt, T. C. Sun, and Nicolas Tserpes.

Göran Högnäs and Arunava Mukherjea

## References

Barnsley, M. F., *Fractals Everywhere*, Academic Press, Orlando (1988).

Crutchfield, J. P. , J. D. Farmer, N. H. Packard, and R. S. Shaw, "Chaos," *Scientific American* **255,** No. 6, 38–49 (1986).

Hansel, G. and D. Perrin, "Codes and Bernoulli partitions," *Math. Systems Theory* **16,** 133–157 (1983).

Högnäs, G. and A. Mukherjea, "Recurrent random walks and invariant measures on semigroups of $n \times n$ matrices," *Math. Zeitschrift* **173,** 69–94 (1980).

Kifer, Y., *Ergodic Theory of Random Transformations*, Birkhäuser, Boston–Basel–Stuttgart (1986).

Ruzsa, I. Z., "Logarithmic density and measures on semigroups," *(Preprint)* (1994).

# Contents

# 1

# Semigroups

## 1.1. Introduction

Chapter 1 contains the basics of semigroups: definitions, elementary concepts, and fundamental examples. We assume some familiarity with standard notions of point-set topology [see Kelley (1955), Mukherjea and Pothoven (1984)]; the algebraic portions of Chapter 1 are however completely self-contained. Without going into any detail whatsoever, it is perhaps prudent to remark at this point that our main interest centers around asymptotics, invariance questions, etc. Our treatment is a reflection of this. We concentrate on algebraic concepts corresponding to such phenomena as absorption, stability, and invariance: zeros, simple semigroups, minimal ideals, maximal subgroups, and so on. We strive to keep digressions at a minimum. Clifford and Preston (1961) offer a wealth of information on all aspects of algebraic semigroups, and this text is recommended to any reader interested in a much more elaborate treatment of this fascinating subject.

Sections 1.1 to 1.5 contain basic material necessary for the development of all subsequent chapters, while Sections 1.6 to 1.8, which deal with more specific applications, can be skipped at first reading.

Arguably the most important notion in mathematics is that of a *mapping* (or *function* or *transformation*). The ultimate goal of the research presented in this book is to describe the long-term behavior of random transformations of some set. Transformations of a set form a semigroup in a natural way. Indeed we see that any semigroup is (algebraically) a transformation semigroup in a canonical way. *Linear transformations* of a vector space form another family of fundamental examples. We devote considerable effort to those semigroups, which incidentally may just as well be viewed as semigroups of *matrices*.

Let $S$ be a set. If $S$ is endowed with an associative binary operation [which we call *multiplication* and denote by a dot $(\cdot)$ or simply by juxtaposition] then $S$

is called a *semigroup*. Strictly speaking the semigroup is the pair $(S, \cdot)$, but the intended operation is usually quite clear from the context. When we are dealing with a specific application we of course use the established notation.

If $s$ is an element of a semigroup $S$ and $A$ and $B$ are subsets of $S$, then we denote by $sA$ the set $\{sa \in S | a \in A\}$ and by $AB$ the set $\{ab \in S | a \in A, b \in B\}$, ($As$ is, of course, defined analogously). Note that products with more than two factors, such as $abc$ and $aBcS$, are well-defined due to the associativity of the multiplication. $aa, aaa, \ldots$ are usually written $a^2, a^3, \ldots$.

A nonempty subset $T$ of $S$ is called a *subsemigroup* if it is stable under multiplication; i.e., if $TT \subset T$. If $T$ is also a group, we of course call it a *subgroup* of $S$. As we see later, it is important in many applications to identify subgroups of a given semigroup $S$.

A subsemigroup $L$ of $S$ is called a *left ideal* if $SL \subset L$; *right* ideals are similarly defined. A nonempty subsemigroup $I$ is a bilateral, or two-sided ideal or just *ideal*, of $S$ if it is both a right and a left ideal: $SI \subset I, IS \subset I$. $S$ is said to be *left (right) simple* if it contains no proper left (right) ideal. Similarly $S$ is *simple* if the only ideal of $S$ is $S$ itself. A left (right) ideal is a *principal* left (right) ideal if it is of the form $\{a\} \cup Sa$ ($\{a\} \cup aS$) for some $a \in S$.

Note that $S$ is left simple if and only if for any given $a, b \in S$, the equation $xa = b$ is soluble. ($Sa$ is a left ideal of $S$ for all $a \in S$. On the other hand, any left ideal $L$ contains a subset of the form $Sa$.)

An element $e \in S$ is called a *left (right) identity element of* $S$ if $es = s$ $(se = s)$ for every $s \in S$. $e$ is a *two-sided identity element of* $S$, or simply *identity of* $S$, if it is both a right and a left identity of $S$. It is easy to see that the identity is unique if it exists.

An element $z$ of $S$ is called a *left (right) zero element of* $S$ if $zs = z$ $(sz = z)$ for all $s \in S$. If $z$ is both a left zero and a right zero of $S$, we simply call it a *zero* of $S$. A semigroup has at most one zero.

The semigroup $S$ is said be *left (right) cancellative* if for any $a \in S$, the equation $ax = ay$ $(xa = ya)$ in $S$ implies $x = y$.

An element $a \in S$ is *idempotent* if $a^2 = a$. Zeros and identities are idempotent. An idempotent is in a trivial fashion the identity element of a subgroup of $S$.

Elements $a$ and $b$ of $S$ are said to *commute* if $ab = ba$. If all elements of $S$ commute with each other, $S$ is called a *commutative* or *abelian* semigroup. In abelian semigroups the operation is often called *addition* and denoted by +; the identity element is denoted by 0 and the inverse of $a$ by $-a$.

To put the preceding concepts into perspective, we now investigate some of their relationships to a group structure. Most textbooks define a *group* as a nonempty set $G$ with an associative binary operation with identity element $e$ and inverses; i.e., for all $a \in G$, there is a $b \in G$ such that $ab = ba = e$, [see, for example, Durbin (1979) or *Encyclopædia Britannica* (1982)]. There is however

a multitude of alternative, seemingly weaker, but in fact equivalent definitions, [see Clifford and Preston (1961), chap. 1].

In our context, Proposition 1.1 is a convenient characterization of a group.

PROPOSITION 1.1. *A semigroup is a group if and only if it is both left and right simple.*

PROOF. A group is clearly both left and right simple.

Conversely, let $S$ be a semigroup that is both left and right simple. For any $a \in S$, the equation $ae = a$ has a solution. On the other hand, $Sa$ is all of $S$; hence $e$ is a right identity of $S$. In the same way we can produce a left identity $f$ that turns out to be equal to $e$. (We have $f = fe = e$.)

It is just as straightforward to obtain the two-sided inverse for an element $a \in S$. Let $b$ and $c$ be solutions of equations $ab = e$ and $ca = e$, respectively. Then $b = eb = (ca)b = c(ab) = ce = c$. Thus $S$ is a group. □

EXAMPLE 1.1. The semigroup of transformations of a set.

Let $X$ be any finite or infinite nonempty set. If $f$ and $g$ are mappings from $X$ to itself, we define as usual the *composition* $f \circ g$ of $f$ and $g$ by $(f \circ g)(x) = f(g(x)), x \in X$. (The domain of the mappings is always understood to be all of $X$.) The composition of mappings is an associative operation; hence the set of all mappings from $X$ into $X$ forms a semigroup with composition as multiplication. This semigroup is called the *full transformation semigroup on X* and denoted by $\mathcal{T}_X$. A more complete treatment of $\mathcal{T}_X$ is given in Clifford and Preston (1961), chap. 2.2.

$\mathcal{T}_X$ has an identity element, the identity mapping $\iota$: $\iota(x) = x, x \in X$. The *constant* mappings are left zeros, since $c \circ f = c$ for all $f$ if $c(x) = x_0, x \in X$ (where $x_0$ is a particular element of $X$). On the other hand, $\mathcal{T}_X$ does not admit any right zeros unless of course $X$ is a singleton.

Define the *range* $R(f)$ of an $f \in \mathcal{T}_X$ to be the set $f(X) \equiv \{f(x) \mid x \in X\}$ (where $\equiv$ means equal by definition). Clearly $R(f \circ g) \subset R(f)$. If $R_0$ is a subset of $X$, then mappings with range inside $R_0$, $\{f \in \mathcal{T}_X | R(f) \subset R_0\}$, form a right ideal of $\mathcal{T}_X$.

The *partition* $\pi(f)$ of $X$ generated by an element $f \in \mathcal{T}_X$ is the equivalence relation on $X$ defined by $x\pi(f)y \iff f(x) = f(y), x, y \in X$. In other words, elements are $\pi(f)$-equivalent if and only if they have the same image under $f$. The $\pi(g)$-equivalence implies $\pi(f \circ g)$-equivalence, $\pi(f \circ g) \supset \pi(g)$. Consequently, if $\pi_0$ is a given equivalence relation on $X$, then those $f \in \mathcal{T}_X$ with $\pi(f) \supset \pi_0$ form a left ideal of $\mathcal{T}_X$.

Define the *rank* of $f$ to be the cardinality of $R(f)$, which we denote by

$|R(f)|$. Note that the cardinality of $R(f)$ is the same as that of the quotient space $X/\pi(f)$ [the number of $\pi(f)$-equivalence classes]. Mappings with a rank no larger than a given cardinality $r$, $\{f \in \mathcal{T}_X |\ |R(f)| \leq r\}$, form a two-sided ideal of the semigroup $\mathcal{T}_X$.

The constant mappings; i.e., mappings of rank 1, form the minimal two-sided ideal of $\mathcal{T}_X$. This subsemigroup is a *left zero semigroup*, since all its elements are left zeros.

Let us now identify the subgroups of $\mathcal{T}_X$. Since any subgroup has an idempotent as its identity element, our first task is to determine idempotents in $\mathcal{T}_X$. Let $e$ be idempotent with range $R$ and partition $\pi$, when $e(x) = x, x \in R$. As before $|R| = |X/\pi|$. This is possible if and only if $R$ is a *complete set of representatives* of the $\pi$-equivalence classes; i.e., every equivalence class contains exactly one element of $R$. In the terminology of Clifford and Preston (1961), $R$ is a *cross section* of $\pi$.

Suppose $f$ belongs to a subgroup $G$ of $\mathcal{T}_X$ and the identity element of $G$ is the idempotent $e$ previously discussed. We can immediately conclude from the relation $e \circ f = f \circ e = f$ that the range of $f$ is $R$ and the partition corresponding to $f$ is $\pi$. Furthermore $f$ is a one-to-one mapping from $R$ to $R$ precisely because $R$ is a cross section of $\pi$. We can then construct a $g$ belonging to $\mathcal{T}_X$ with the following properties: $g$ has range $R$ and partition $\pi$, and the restriction of $g$ to $R$ is the inverse of the restriction of $f$ to $R$. It is then clear that $g \circ f = f \circ g = e$, in other words, the inverse of $f$ in the subgroup $G$ is $g$. Our result is thus the following, [see Clifford and Preston (1961), theorem 2.10]: *$f$ belongs to a subgroup of $\mathcal{T}_X$ if and only if $R(f)$ is a complete set of representatives of $\pi(f)$. The sets $\{f | R(f) = R, \pi(f) = \pi\}$ where $|R| = |X/\pi|$ are groups if and only if $R$ is a complete set of representatives of $\pi$.*

If there is an $f$ whose range is *not* a cross section of $\pi(f)$ then $f \circ f$ has a smaller range than $f$: $R(f \circ f)$ is a proper subset of $R(f)$. Clearly, such an $f$ cannot belong to a group.

The preceding discussion holds almost verbatim for any transformation semigroup on $X$; i.e., any subsemigroup $S$ of $\mathcal{T}_X$. The conditions are necessary only in the general case. For example, $f \in S$ can belong to a subgroup of $S$ only if $R(f)$ is a cross section of $\pi(f)$. For a transformation semigroup $S$ on a *finite* set $X$, we have the converse: If elements of a subsemigroup of $S$ have common range $R$ and partition $\pi$, where $R$ is a cross section of $\pi$, then it is a group. (Each element $f$ is of finite order; i.e., some power of $f$ equals the identity mapping $e$ on $R$ with the given partition $\pi$.)

Let $X$ be countably infinite and $G$ a subgroup of infinite rank of $\mathcal{T}_X$. If $e$ is the identity element of $G$, then we can construct a mapping $\alpha$ with the properties that $\alpha$ restricted to the range of $e$ is a bijection onto all of $X$ (practically the definition of an infinite subset of $X$) and $\alpha$ has the same partition as $e$. If we denote the inverse of $\alpha$ restricted to the range of $e$ by $\beta$, then $\beta$ is injective, and

has the same range as $e$. Hence $\alpha \circ \beta = \iota$ (the identity mapping on $X$) and $\beta \circ \alpha = e$. For any $g \in G$, $\alpha \circ g \circ \beta$ is a bijection on $X$. Conversely elements of $G$ can be written $\beta \circ h \circ \alpha$, where the $h$s are bijections on $X$.

## 1.2. Homomorphisms, Quotients, and Products

A mapping $\phi$ between two semigroups $(S, \cdot)$ and $(T, *)$ is called a *semigroup homomorphism (antihomomorphism)* if

$$\phi(a \cdot b) = \phi(a) * \phi(b) \ \ (\phi(a \cdot b) = \phi(b) * \phi(a)), \ \ a, b \in S.$$

The $\phi$ is said to be an *semigroup isomorphism (antiisomorphism)* if it is bijective (i.e., onto and one-to-one) as well. We however usually suppress the explicit reference to semigroups when it is clear from the context that we are dealing with a semigroup structure. The $S$ and $T$ are *isomorphic* (as *semigroups*) if there exists a semigroup isomorphism between them. Using the antiisomorphism $a \cdot b \mapsto b * a$ we can always convert an antihomomorphism to a homomorphism if need be.

As an example, consider the set of *left translations* on a given semigroup $S$: $\lambda_a, a \in S$, where $\lambda_a$ is defined by $\lambda_a(x) = ax, x \in S$. Clearly $\lambda_{ab} = \lambda_a \circ \lambda_b$. The $\lambda$ is thus a homomorphism from $S$ to the full transformation semigroup on $S$, $\mathcal{T}_S$. Right translations define in a similar way a semigroup antihomomorphism from $S$ to $\mathcal{T}_S$. If in addition $ax = bx$ for all $x \in S$ implies $a = b$ (the left translations *act effectively on S*), then $\lambda$ is injective and *S is isomorphic to a subsemigroup of* $\mathcal{T}_S$. In particular this is the case when $S$ has a right identity element.

If we extend the idea of left translations slightly, we obtain the useful result that any semigroup *S is isomorphic to some transformation semigroup.* We need only take $X = S \cup 1$ and define $a1 = a, a \in S$. Left translations $\lambda_a, a \in S$ on $X$ define an injective homomorphism from $S$ into $\mathcal{T}_X$.

An equivalence relation $\rho$ on a semigroup $S$ compatible with the multiplication is called a *congruence* on $S$. More formally $\rho$ is a congruence if for all $a, b, s \in S$, $a\rho b$ implies $as\rho bs$ and $sa\rho sb$. If $\rho$ is a congruence on $S$, then the multiplication of equivalence classes in the natural way is a a well-defined operation on $S/\rho$. (For $a, b \in S$, we take $[a][b] = [ab]$, where $[a]$ is the equivalence class containing $a$). The semigroup thus obtained is called the *quotient* or *factor semigroup* of $S \bmod \rho$.

The discussion in the preceding paragraphs shows that if $\rho$ is the congruence on $S$ defined by

$$a\rho b \iff ax = bx \text{ for all } x \in S$$

then $S/\rho$ is isomorphic to a subsemigroup of $\mathcal{T}_S$.

Definition 1.1 presents another useful congruence in semigroup theory.

DEFINITION 1.1. Let $I$ be an ideal of $S$. If we define a relation $\rho$ by

$$a\rho b \iff a = b \text{ or else both } a \text{ and } b \in I$$

then $\rho$ is a congruence on $S$. The corresponding factor semigroup is usually written $S/I$, and it is called the *Rees factor* (or *quotient*) *semigroup* of $S \bmod I$. The intuitive idea behind the Rees factor semigroup is to lump all elements of $I$ together into a single zero.

Take a semigroup $S$ and view it as a transformation semigroup on $X$. For $f \in S$, define a *matrix* $B_f(x, y)$ indexed by $X$ according to the following prescription: $B_f(x, y) = \delta_{f(y)}x$, i.e., $B_f(x, y) = 1$ if $x = f(y)$ and 0 otherwise, $x, y \in X$. Multiplying the matrices according to the usual rules of matrix multiplication, we see that indeed $B_{f \circ g} = B_f B_g$, [see Darling and Mukherjea (1988) where their matrices $A_f$ have the antihomomorphism property instead]. Hence any semigroup can be described as a semigroup of 0-1 matrices, *transformation matrices*, with exactly one 1 in each column. If there is exactly one 1 in each row as well, we obtain the familiar *permutation* matrices corresponding to bijections on the set $X$.

For semigroups $(S, \cdot)$ and $(T, *)$ we obtain a new structure on their cartesian product by the rule

$$(s, t) \star (s', t') = (s \cdot s', t * t').$$

The resulting semigroup $(S \times T, \star)$ is called the *direct product* of $(S, \cdot)$ and $(T, *)$. Direct products with several factors are defined analogously.

Let $G$ be a group and $E$ a right zero semigroup (where $ee' = e'$, $e, e' \in E$). Consider the direct product of $G$ and $E$. Multiplication in $G \times E$ is given by $(g, e)(g', e') = (gg', e')$.

The usefulness of this structure is due to the fact that any right group has this representation.

DEFINITION 1.2. A *right group* (*left group*) is a semigroup that is right simple (left simple) and left cancellative (right cancellative).

Alternative characterizations are given in Proposition 1.2.

PROPOSITION 1.2. *For a semigroup $S$, the following statements are equivalent:*

*(i) $S$ is a right group;*

*(ii) For any $a, b \in S$ the equation $ax = b$ has one and only one solution;*

*(iii)* $S$ *is right simple and contains an idempotent;*

*(iv)* $S$ *is isomorphic to the direct product of a group* $G$ *and a right zero semigroup* $E$.

PROOF. The equivalence of (i) and (ii) follows immediately from the definitions and the characterization of right simplicity in Section 1.1.

Let $S$ satisfy (ii). Then for any $a \in S$ the equation $ax = a$ has a solution $e$, say. We have $aee = ae$; by left cancellativity, $ee = e$. Thus (ii) implies (iii).

Assume (iii) and let $ax = ay$ for some $a, x, y \in S$. There is an idempotent $e$ in $S$, $e$ is a left identity of $S = eS$. There is a $b \in S$ such that $ab = e$. $(ba)(ba) = b(ab)a = bea = ba$, so $ba$ is an idempotent too and consequently a left identity of $S$. We finally obtain $ax = ay \Longrightarrow bax = bay \Longrightarrow x = y$; i.e., (i) holds. The equation $(g, e)(x, y) = (g', e')$ in the direct product $G \times E$ has the unique solution $(x, y) = (g^{-1}g', e')$. This shows that (iv) implies (ii).

Suppose now that $S$ satisfies (iii) and equivalently (i) and (ii). Let $E$ be the set of idempotents in $S$. We saw that any idempotent is a left identity, so $ee' = e', e, e' \in E$. In other words, $E$ is a right zero (sub)semigroup (of $S$). Take an $e \in E$. We show next that $Se$ is a subgroup of $S$. $Se$ is right simple: $seSe = Se$. $Se$ is also left simple. To see this, take an element $se \in Se$ and let $t \in S$ be such that $set = e$. Then $se(te) = (set)e = ee = e$. For an arbitrary $ue \in Se$ the equation $(se)(xe) = ue$ can be solved, namely by $x = teu \in S$. By Proposition 1.1, $Se$ is a group. Take a particular idempotent $e_0$ and let $Se_0 = G$.

Consider the map $(g, e) \mapsto ge, g \in G, e \in E$. Call it $\phi$. We prove that $\phi$ is the desired isomorphism.

The map $\phi$ is a homomorphism: For $g, h \in G$ and $e, f \in E$, $gehf = ghf$ because each idempotent is a left identity.

$\phi$ is injective. To see this, let $ge = hf$. Then $g = ge_0 = g(ee_0) = (ge)e_0 = (hf)e_0 = h(fe_0) = he_0 = h$. Left cancellativity then yields $e = f$.

$\phi$ is surjective, since any $a \in S$ can be written in the form $ae$ for some idempotent $e$ [see the proof of (ii) $\Longrightarrow$ (iii) above]. $e_0$ is a left identity of $S$, so $a = a(e_0e) = (ae_0)e$. □

To understand right groups concretely, let us look at the full transformation semigroup $\mathcal{T}_X$ and its subsemigroups. The possible idempotents and subgroups are characterized in Section 1.1. Note that subgroups consist of bijections (permutations) of the set $R$. It is not difficult to see that subgroups are *isomorphic to a permutation group on* $R$ (i.e., a subgroup of the symmetric group $\mathcal{G}_R$ on $R$ consisting of all the bijections from $R$ to $R$).

It is evident that a right group $S \subset \mathcal{T}_X$ consists of mappings with a common range $R_0$; otherwise, it could not possibly be right simple. It follows from the assumption that $S$ is closed under multiplication that the common range $R_0$ is

a cross section of all partitions generated by elements of $S$. Indeed we obtain Proposition 1.3 in the case of a finite $X$.

PROPOSITION 1.3. *A subsemigroup $S$ of $\mathcal{T}_X$ is a right group if and only if the elements of $S$ have common range.*

PROOF. The only if statement is immediate, so let us concentrate on the sufficiency, the if statement. Any $f \in S$ is bijective on the common range, so left cancellativity follows. Since $X$ is finite, the powers $f^n, n = 1, 2, 3 \ldots$ form a group whose identity is a left identity of $S$. The equation $f \circ g = h$ thus has a solution $g = f^{r-1} \circ h$ (where $r$ is the order of the group generated by $f$). In other words, $S$ is right simple and hence a right group. □

To see how the concepts work for countably infinite $X$, let us return to the last paragraph of Example 1.1. If $H$ is a subgroup of the symmetric group $\mathcal{G}_X$, then $\beta \circ H \circ \alpha$ is an infinite-rank subgroup of $\mathcal{T}_X$. Conversely any such subgroup is isomorphic to $\beta \circ H \circ \alpha$ for some $H, \alpha$ and $\beta$. Right groups are then obtained by varying the partition in the construction of $\alpha$: Any right group of infinite rank is isomorphic to $\beta \circ H \circ E$, where $E$ is some set of mappings constructed exactly as $\alpha$ but with the partition varying; of course the fundamental property of the range as a cross section of the partition must be maintained. In particular for an $\alpha' \in E$, $\alpha' \circ \beta = \iota$ and $\beta \circ \alpha'$ is the identity of some group of mappings with the same range as $e$.

DEFINITION 1.3. We conclude this section by introducing another important product structure, the *Rees product.* Let $E$ be a left zero semigroup, $F$ a right zero semigroup, $G$ a group, and $\phi$ a function from $F \times E$ to $G$. Define a multiplication on $E \times G \times F$ by

$$(e, g, f)(e', g', f') = (e, g\phi(f, e')g', f').$$

Note that this product is direct if the *sandwich function* $\phi$ maps everything onto the identity element of the group $G$. Such a $\phi$ is termed *trivial.*

We emphasize at this point that $\phi$ is a completely arbitrary function from $F \times E$ to $G$. Different choices of $\phi$ may produce isomorphic semigroups. If, for example, $\phi$ maps everything onto a a constant $c \in G$, then the resulting Rees product is isomorphic to the direct product of $E$, $G$, and $F$. We will return briefly to this question in Section 1.4 (Proposition 1.13).

REMARK 1.1. The cylinder subsets of the form $\{e\} \times G \times \{f\}$, called *cells,* are all groups isomorphic to $G$. The identity of such a group is the element

$(e, (\phi(f,e))^{-1}, f)$. Any subsemigroup of the form $\{e\} \times G \times B$ (where $B \subset F$) of the Rees product is a right group. This fact is an immediate consequence of Definition 1.2.

## 1.3. Semigroups with Zero

Recall that an element $z$ of a semigroup $S$ is a *zero* if $sz = zs = z$ for all $s \in S$. A zero is unique if it exists. We henceforth adhere to the common convention and denote a zero by 0.

The notions of (right, left) simplicity are trivial in the presence of a zero 0. For instance $S$ is simple if and only if $S = \{0\}$. For semigroups with zero, it is therefore useful for many purposes to restrict some of the definitions to *nonzero* elements of $S$. It is often practical to have a special notation for these elements; $A^*$ is the generic notation for nonzero elements of the set $A$

$$A^* \equiv A \setminus \{0\} \equiv \{a \in A | a \neq 0\}.$$

An ideal $I \neq \{0\}$ of a semigroup $S$ is said to be 0-*minimal* if $\{0\}$ is the only ideal of $S$ properly contained in $I$. A 0-minimal left (right) ideal is defined analogously.

A semigroup $S$ is called *right (left)* 0-*simple* if $S^2 \neq \{0\}$ and its only right (left) ideals are 0 and $S$ itself. $S$ is called 0-simple if $S^2 \neq \{0\}$ and 0 is the only proper two-sided ideal of $S$.

A semigroup $S$ with the property that all products are 0, $S^2 = \{0\}$, is called a *null* semigroup. Such a semigroup obviously satisfies the second condition of the preceding definitions.

Elements (nonzero elements) $a$ and $b$ of a semigroup with 0 are called *divisors of zero (proper divisors of zero)* if $ab = 0$.

Propositions 1.1 and 1.2 have their counterparts for semigroups with 0 which follow immediately from the original Propositions and Lemma 1.6.

PROPOSITION 1.4. *A semigroup with* 0 *is a group with* 0 *if and only if it is both left and right* 0*-simple.*

PROPOSITION 1.5. *For a semigroup* $S$ *with* 0 *the following statements are equivalent:*

*(i)* $S^*$ *is a right group;*

*(ii) for any* $a \in S^*, b \in S$ *the equation* $ax = b$ *has one and only one solution;*