CAMBRIDGE STUDIES IN
ADVANCED MATHEMATICS 105

ADDITIVE COMBINATORICS

Additive combinatorics is the theory of counting additive structures in sets. This theory has seen exciting developments and dramatic changes in direction in recent years, thanks to its connections with areas such as number theory, ergodic theory and graph theory. This graduate level textbook will allow students and researchers easy entry into this fascinating field. Here, for the first time, the authors bring together, in a self-contained and systematic manner, the many different tools and ideas that are used in the modern theory, presenting them in an accessible, coherent, and intuitively clear manner, and providing immediate applications to problems in additive combinatorics. The power of these tools is well demonstrated in the presentation of recent advances such as the Green-Tao theorem on arithmetic progressions and Erdős distance problems, and the developing field of sum-product estimates. The text is supplemented by a large number of exercises and new material.

Terence Tao is a professor in the Department of Mathematics at the University of California, Los Angeles.

Van Vu is a professor in the Department of Mathematics at Rutgers University, New Jersey.

CAMBRIDGE STUDIES IN ADVANCED MATHEMATICS

Au the title, listed below can be obtained from good booksellers or from Cambridge University Press for a complete listing visit www.cambridge.org/uk/series/&Series.asp?code=CSAM.

49 R. Stanley *Enumerative combinatorics I*

50 I. Porteous *Clifford algebras and the classical groups*

51 M. Audin *Spinning tops*

52 V. Jurdjevic *Geometric control theory*

53 H. Volklein *Groups as Galois groups*

54 J. Le Potier *Lectures on vector bundles*

55 D. Bump *Automorphic forms and representations*

56 G. Laumon *Cohomology of Drinfeld modular varieties II*

57 D.M. Clark & B.A. Davey *Natural dualities for the working algebraist*

58 J. McCleary *A user's guide to spectral sequences II*

59 P. Taylor *Practical foundations of mathematics*

60 M.P. Brodmann & R.Y. Sharp *Local cohomology*

61 J.D. Dixon *et al. Analytic pro-P groups*

62 R. Stanley *Enumerative combinatorics II*

63 R.M. Dudley *Uniform central limit theorems*

64 J. Jost & X. Li-Jost *Calculus of variations*

65 A.J. Berrick & M.E. Keating *An introduction to rings and modules*

66 S. Morosawa *Holomorphic dynamics*

67 A.J. Berrick & M.E. Keating *Categories and modules with K-theory in view*

68 K. Sato *Levy processes and infinitely divisible distributions*

69 H. Hida *Modular forms and Galois cohomology*

70 R. Iorio & V. Iorio *Fourier analysis and partial differential equations*

71 R. Blei *Analysis in integer and fractional dimensions*

72 F. Borceaux & G. Janelidze *Galois theories*

73 B. Bollobas *Random graphs*

74 R.M. Dudley *Real analysis and probability*

75 T. Sheil-Small *Complex polynomials*

76 C. Voisin *Hodge theory and complex algebraic geometry I*

77 C. Voisin *Hodge theory and complex algebraic geometry II*

78 V. Paulsen *Completely bounded maps and operator algebras*

79 F. Gesztesy & H. Holden *Soliton Equations and their Algebro-Geometric Solutions Volume 1*

81 Shigeru Mukai *An Introduction to Invariants and Moduli*

82 G. Tourlakis *Lectures in logic and set theory I*

83 G. Tourlakis *Lectures in logic and set theory II*

84 R.A. Bailey *Association Schemes*

85 James Carlson, Stefan Müller-Stach, & Chris Peters *Period Mappings and Period Domains*

86 J.J. Duistermaat & J.A.C. Kolk *Multidimensional Real Analysis I*

87 J.J. Duistermaat & J.A.C. Kolk *Multidimensional Real Analysis II*

89 M. Golumbic & A.N. Trenk *Tolerance Graphs*

90 L.H. Harper *Global Methods for Combinatorial Isoperimetric Problems*

91 I. Moerdijk & J. Mrcun *Introduction to Foliations and Lie Groupoids*

92 János Kollár, Karen E. Smith, & Alessio Corti *Rational and Nearly Rational Varieties*

93 David Applebaum *Lévy Processes and Stochastic Calculus*

95 Martin Schechter *An Introduction to Nonlinear Analysis*

Additive Combinatorics

TERENCE TAO, VAN VU

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521853866

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN-13 978-0-521-85386-6 hardback
ISBN-10 0-521-85386-9 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

To our families

Cambridge University Press
0521853869 - Additive Combinatorics
Terence Tao and Van Vu
Frontmatter
More information

Contents

Prologue

This book arose out of lecture notes developed by us while teaching courses on additive combinatorics at the University of California, Los Angeles and the University of California, San Diego. Additive combinatorics is currently a highly active area of research for several reasons, for example its many applications to additive number theory. One remarkable feature of the field is the use of tools from many diverse fields of mathematics, including elementary combinatorics, harmonic analysis, convex geometry, incidence geometry, graph theory, probability, algebraic geometry, and ergodic theory; this wealth of perspectives makes additive combinatorics a rich, fascinating, and multi-faceted subject. There are still many major problems left in the field, and it seems likely that many of these will require a combination of tools from several of the areas mentioned above in order to solve them.

The main purpose of this book is to gather all these diverse tools in one location, present them in a self-contained and introductory manner, and illustrate their application to problems in additive combinatorics. Many aspects of this material have already been covered in other papers and texts (and in particular several earlier books [168], [257], [116] have focused on some of the aspects of additive combinatorics), but this book attempts to present as many perspectives and techniques as possible in a unified setting.

Additive combinatorics is largely concerned with the additive structure[1] of sets. To clarify what we mean by "additive structure", let us introduce the following definitions.

Definition 0.1 An *additive group* is any abelian group Z with group operation $+$. Note that we can define a multiplication operation $nx \in Z$ whenever $n \in \mathbf{Z}$ and

[1] We will also occasionally consider the multiplicative structure of sets as well; we will refer to the combined study of such structures as *arithmetic combinatorics*.

$x \in Z$ in the usual manner: thus $3x = x + x + x$, $-2x = -x - x$, etc. An *additive set* is a pair (A, Z), where Z is an additive group, and A is a finite non-empty subset of Z. We often abbreviate an additive set (A, Z) simply as A, and refer to Z as the *ambient group* of the additive set. If A, B are additive sets in Z, we define the *sum set*

$$A + B := \{a + b : a \in A,\ b \in B\}$$

and *difference set*

$$A - B := \{a - b : a \in A,\ b \in B\}.$$

Also, we define the *iterated sumset* kA for $k \in \mathbf{Z}^+$ by

$$kA := \{a_1 + \cdots + a_k : a_1, \ldots, a_k \in A\}.$$

We caution that the sumset kA is usually distinct from the dilation $k \cdot A$ of A, defined by

$$k \cdot A := \{ka : a \in A\}.$$

For us, typical examples of additive groups Z will be the integers $\mathbf{Z}$, a cyclic group $\mathbf{Z}_N$, a Euclidean space $\mathbf{R}^n$, or a finite field geometry F_p^n. As the notation suggests, we will eventually be viewing additive sets as "intrinsic" objects, which can be embedded inside any number of different ambient groups; this is somewhat similar to how a manifold can be thought of intrinsically, or alternatively can be embedded into an ambient space. To make these ideas rigorous we will need to develop the theory of *Freiman homomorphisms*, but we will defer this to Section 5.3.

Additive sets may have a large or small amount of additive structure. A good example of a set with little additive structure would be a randomly chosen subset A of a finite additive group Z with some fixed cardinality. At the other extreme, examples of sets with very strong additive structure would include arithmetic progressions

$$a + [0, N) \cdot r := \{a, a + r, \ldots, a + (N - 1)r\}$$

where $a, r \in Z$ and $N \in \mathbf{Z}^+$; or d-dimensional generalized arithmetic progressions

$$a + [0, N) \cdot v := \{a + n_1 v_1 + \cdots + n_d v_d : 0 \le n_j < N_j \text{ for all } 1 \le j \le d\}$$

where $a \in Z$, $v = (v_1, \ldots, v_d) \in Z^d$, and $N = (N_1, \ldots, N_d) \in (\mathbf{Z}^+)^d$; or d-dimensional cubes

$$a + \{0, 1\}^d \cdot v = \{a + \epsilon_1 v_1 + \cdots + \epsilon_d v_d : \epsilon_1, \ldots, \epsilon_d \in \{0, 1\}\};$$

or the subset sums $FS(A) := \{\sum_{a \in B} a : B \subseteq A\}$ of a finite set A.

A fundamental task in this subject is to give some quantitative measures of additive structure in a set, and then investigate to what extent these measures are equivalent to each other. For example, one could try to quantify each of the following informal statements as being some version of the assertion "A has additive structure":

- $A + A$ is small;
- $A - A$ is small;
- $A - A$ can be covered by a small number of translates of A;
- kA is small for any fixed k;
- there are many quadruples $(a_1, a_2, a_3, a_4) \in A \times A \times A \times A$ such that $a_1 + a_2 = a_3 + a_4$;
- there are many quadruples $(a_1, a_2, a_3, a_4) \in A \times A \times A \times A$ such that $a_1 - a_2 = a_3 - a_4$;
- the convolution $1_A * 1_A$ is highly concentrated;
- the subset sums $FS(A) := \{\sum_{a \in B} a : B \subseteq A\}$ have high multiplicity;
- the Fourier transform $\widehat{1_A}$ is highly concentrated;
- the Fourier transform $\widehat{1_A}$ is highly concentrated in a cube;
- A has a large intersection with a generalized arithmetic progression, of size comparable to A;
- A is contained in a generalized arithmetic progression, of size comparable to A;
- A (or perhaps $A - A$, or $2A - 2A$) contains a large generalized arithmetic progression.

The reader is invited to investigate to what extent these informal statements are true for sets such as progressions and cubes, and false for sets such as random sets. As it turns out, once one makes the above assertions more quantitative, there are a number of deep and important equivalences between them; indeed, to oversimplify tremendously, all of the above criteria for additive structure are "essentially" equivalent. There is also a similar heuristic to quantify what it would mean for two additive sets A, B of comparable size to have a large amount of "shared additive structure" (e.g. A and B are progressions with the same step size v); we invite the reader to devise analogs of the above criteria to capture this concept.

Making the above heuristics precise and rigorous will require some work, and in fact will occupy large parts of Chapters 2, 3, 4, 5, 6. In deriving these basic tools of the field, we shall need to develop and combine techniques from elementary combinatorics, additive geometry, harmonic analysis, and graph theory; many of these methods are of independent interest in their own right, and so we have devoted some space to treating them in detail.

Of course, a "typical" additive set will most likely behave like a random additive set, which one expects to have very little additive structure. Nevertheless, it is a deep and surprising fact that as long as an additive set is dense enough in its ambient group, it will always have *some* level of additive structure. The most famous example of this principle is *Szemerédi's theorem*, which asserts that every subset of the integers of positive upper density will contain arbitrarily long arithmetic progressions; we shall devote all of Chapter 11 to this beautiful and important theorem. A variant of this fact is the very recent *Green–Tao theorem*, which asserts that every subset of the prime numbers of positive upper *relative* density also contains arbitrarily long arithmetic progressions; in particular, the primes themselves have this property. If one starts with an even sparser set A than the primes, then it is not yet known whether A will necessarily contain long progressions; however, if one forms sum sets such as $A + A$, $A + A + A$, $2A - 2A$, $FS(A)$ then these sets contain extraordinarily long arithmetic progressions (see in particular Section 4.7 and Chapter 12). This basic principle – that sumsets have much more additive structure than general sets – is closely connected to the equivalences between the various types of additive structure mentioned previously; indeed results of the former type can be used to deduce results of the latter type, and conversely.

We now describe some other topics covered in this text. In Chapter 1 we recall the simple yet powerful *probabilistic method*, which is very useful in additive combinatorics for constructing sets with certain desirable properties (e.g. thin additive bases of the integers), and provides an important conceptual framework that complements more classical deterministic approaches to such constructions. In Chapter 6 we present some ways in which graph theory interacts with additive combinatorics, for instance in the theory of sum-free sets, or via Ramsey theory. Graph theory is also decisive in establishing two important results in the theory of sum sets, the Balog–Szemerédi–Gowers theorem and the Plünnecke inequalities. Two other important tools from graph theory, namely the crossing number inequality and the Szemerédi regularity lemma, will also be covered in Chapter 8 and Sections 10.6, 11.6 respectively. In Chapter 7 we view sum sets from the perspective of random walks, and give some classical and recent results concerning the distribution of these sum sets, and in particular recent applications to random matrices. Last, but not least, in Chapter 9 we describe some algebraic methods, notably the combinatorial Nullstellensatz and Chevalley–Waring type methods, which have led to several deep arithmetical results (often with very sharp bounds) not obtainable by other means.

Acknowledgements

The authors would like to thank Shimon Brooks, Robin Chapman, Michael Cowling, Andrew Granville, Ben Green, Timothy Gowers, Harald Helfgott, Martin Klazar, Mariah Hamel, Vsevolod Lev, Roy Meshulam, Melvyn Nathanson, Imre Ruzsa, Roman Sasyk, and Benny Sudakov for helpful comments and corrections,

and to the Australian National University and the University of Edinburgh for their hospitality while portions of this book were being written. Parts of this work were inspired by the lecture notes of Ben Green [144], the expository article of Imre Ruzsa [297], and the book by Melvyn Nathanson [257]. TT is also particularly indebted to Roman Sasyk and Hillel Furstenberg for explaining the ergodic theory proof of Szemerédi's theorem. VV would like to thank Endre Szemerédi for many useful discussions on mathematics and other aspects of life. Last, and most importantly, the authors thank their wives, Laura and Huong, without whom this book would not be finished.

General notation

The following general notational conventions will be used throughout the book.

Sets and functions

For any set A, we use

$$A^d := A \times \cdots \times A = \{(a_1, \ldots, a_d) : a_1, \ldots, a_d \in A\}$$

to denote the Cartesian product of d copies of A: thus for instance $\mathbf{Z}^d$ is the d-dimensional integer lattice. We shall occasionally denote A^d by $A^{\oplus d}$, in order to distinguish this Cartesian product from the d-fold product set $A^{\cdot d} = A \cdot \ldots \cdot A$ of A, or the d-fold powers $A^\wedge d := \{a^d : a \in A\}$ of A.

If A, B are sets, we use $A \backslash B := \{a \in A : a \notin B\}$ to denote the set-theoretic difference of A and B; and B^A to denote the space of functions $f : A \to B$ from A to B. We also use $2^A := \{B : B \subset A\}$ to denote the power set of A. We use $|A|$ to denote the cardinality of A. (We shall also use $|x|$ to denote the magnitude of a real or complex number x, and $|v| = \sqrt{v_1^2 + \cdots + v_d^2}$ to denote the magnitude of a vector $v = (v_1, \ldots, v_d)$ in a Euclidean space $\mathbf{R}^d$. The meaning of the absolute value signs should be clear from context in all cases.)

If $A \subset Z$, we use $1_A : Z \to \{0, 1\}$ to denote the indicator function of A: thus $1_A(x) = 1$ when $x \in A$ and $1_A(x) = 0$ otherwise. Similarly if P is a property, we let $\mathbf{I}(P)$ denote the quantity 1 if P holds and 0 otherwise; thus for instance $1_A(x) = \mathbf{I}(x \in A)$.

We use $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ to denote the number of k-element subsets of an n-element set. In particular we have the natural convention that $\binom{n}{k} = 0$ if $k > n$ or $k < 0$.

Number systems

We shall rely frequently on the integers $\mathbf{Z}$, the positive integers $\mathbf{Z}^+ := \{1, 2, \ldots\}$, the natural numbers $\mathbf{N} := \mathbf{Z}_{\geq 0} = \{0, 1, \ldots\}$, the reals $\mathbf{R}$, the positive reals

$\mathbf{R}^+ := \{x \in \mathbf{R} : x > 0\}$, the non-negative reals $\mathbf{R}_{\geq 0} := \{x \in \mathbf{R} : x \geq 0\}$, and the complex numbers $\mathbf{C}$, as well as the circle group $\mathbf{R}/\mathbf{Z} := \{x + \mathbf{Z} : x \in \mathbf{R}\}$.

For any natural number $N \in \mathbf{N}$, we use $\mathbf{Z}_N := \mathbf{Z}/N\mathbf{Z}$ to denote the cyclic group of order N, and use $n \mapsto n \mod N$ to denote the canonical projection from $\mathbf{Z}$ to $\mathbf{Z}_N$. If q is a prime power, we use F_q to denote the finite field of order q (see Section 9.4). In particular if p is a prime then F_p is identifiable with $\mathbf{Z}_p$.

If x is a real number, we use $\lfloor x \rfloor$ to denote the greatest integer less than or equal to x.

Landau asymptotic notation

Let n be a positive variable (usually taking values on $\mathbf{N}$, $\mathbf{Z}^+$, $\mathbf{R}_{\geq 0}$, or $\mathbf{R}^+$, and often assumed to be large) and let $f(n)$ and $g(n)$ be real-valued functions of n.

- $g(n) = O(f(n))$ means that f is non-negative, and there is a positive constant C such that $|g(n)| \leq Cf(n)$ for all n.
- $g(n) = \Omega(f(n))$ means that f, g are non-negative, and there is a positive constant c such that $g(n) \geq cf(n)$ for all sufficiently large n.
- $g(n) = \Theta(f(n))$ means that f, g are non-negative and both $g(n) = O(f(n))$ and $g(n) = \Omega(f(n))$ hold; that is, there are positive constants c and C such that $cf(n) \geq g(n) \geq Cf(n)$ for all n.
- $g(n) = o_{n\to\infty}(f(n))$ means that f is non-negative and $g(n) = O(a(n)f(n))$ for some $a(n)$ which tends to zero as $n \to \infty$; if f is strictly positive, this is equivalent to $\lim_{n\to\infty} g(n)/f(n) = 0$.
- $g(n) = \omega_{n\to\infty}(f(n))$ means that f, g are non-negative and $f(n) = o_{n\to\infty}(g(n))$.

In most cases the asymptotic variable n will be clear from context, and we shall simply write $o_{n\to\infty}(f(n))$ as $o(f(n))$, and similarly write $\omega_{n\to\infty}(f(n))$ as $\omega(f(n))$. In some cases the constants c, C and the decaying function $a(n)$ will depend on some other parameters, in which case we indicate this by subscripts. Thus for instance $g(n) = O_k(f(n))$ would mean that $g(n) \leq C_k f(n)$ for all n, where C_k depends on the parameter k; similarly, $g(n) = o_{n\to\infty;k}(f(n))$ would mean that $g(n) = O(a_k(n)f(n))$ for some $a_k(n)$ which tends to zero as $n \to \infty$ for each fixed k.

The notation $g(n) = \tilde{O}(f(n))$ has been used widely in the combinatorics and theoretical computer science community in recent years; $g(n) = \tilde{O}(f(n))$ means that there is a constant c such that $g(n) \leq f(n)\log^c n$ for all sufficiently large n. We can define, in a similar manner, $\tilde{\Omega}$ and $\tilde{\Theta}$, though this notation will only be used occasionally here. Here and throughout the rest of the book, log shall denote the natural logarithm unless specified by subscripts, thus $\log_x y = \frac{\log y}{\log x}$.

Progressions

We have already encountered the concept of a generalized arithmetic progression. We now make this concept more precise.

Definition 0.2 (Progressions) For any integers $a \leq b$, we let $[a, b]$ denote the discrete closed interval $[a, b] := \{n \in \mathbf{Z} : a \leq n \leq b\}$; similarly define the half-open discrete interval $[a, b)$, etc. More generally, if $a = (a_1, \ldots, a_d)$ and $b = (b_1, \ldots, b_d)$ are elements of $\mathbf{Z}^d$ such that $a_j \leq b_j$, we define the *discrete box*

$$[a, b] := \{(n_1, \ldots, n_d) \in \mathbf{Z}^d : a_j \leq n_j \leq b_j \text{ for all } 1 \leq j \leq d\},$$

and similarly

$$[a, b) := \{(n_1, \ldots, n_d) \in \mathbf{Z}^d : a_j \leq n_j < b_j \text{ for all } 1 \leq j \leq d\},$$

etc. If Z is an additive group, we define a *generalized arithmetic progression* (or just *progression* for short) in Z to be any set[1] of the form $P = a + [0, N] \cdot v$, where $a \in Z$, $N = (N_1, \ldots, N_d)$ is a tuple, $[0, N] \subset \mathbf{Z}^d$ is a discrete box, $v = (v_1, \ldots, v_d) \in Z^d$, the map $\cdot : \mathbf{Z}^d \times Z^d \to Z$ is the dot product

$$(n_1, \ldots, n_d) \cdot (v_1, \ldots, v_d) := n_1 v_1 + \cdots + n_d v_d,$$

and $[0, N] \cdot v := \{n \cdot v : n \in [0, N]\}$. In other words,

$$P = \{a + n_1 v_1 + \cdots + n_d v_d : 0 \leq n_j \leq N_j \text{ for all } 1 \leq j \leq d\}.$$

We call a the *base point* of P, $v = (v_1, \ldots, v_d)$ the *basis vectors* of P, N the *dimension* of P, d the *dimension* or *rank* of P, and $\mathrm{vol}(P) := |[0, N]| = \prod_{j=1}^{d}(N_j + 1)$ the *volume* of P. We say that the progression P is *proper* if the map $n \mapsto n \cdot v$ is injective on $[0, N]$, or equivalently if the cardinality of P is equal to its volume (as opposed to being strictly smaller than the volume, which can occur if the basis vectors are linearly dependent over $\mathbf{Z}$). We say that P is *symmetric* if $-P = P$; for instance $[-N, N] \cdot v = -N \cdot v + [0, 2N] \cdot v$ is a symmetric progression.

Other notation

There are a number of other definitions that we shall introduce at appropriate junctures and which will be used in more than one chapter of the book. These include the probabilistic notation (such as **E**(), **P**(), **I**(), **Var**(), **Cov**()) that we introduce

[1] Strictly speaking, this is an abuse of notation; the arithmetic progression should really be the sextuple (P, d, N, a, v, Z), because the set P alone does not always uniquely determine the base point, step, ambient space or even length (if the progression is improper) of the progression P. However, as it would be cumbersome continually to use this sextuple, we shall usually just P to denote the progression.

Cambridge University Press
0521853869 - Additive Combinatorics
Terence Tao and Van Vu
Frontmatter
More information

at the start of Chapter 1, and measures of additive structure such as the doubling constant $\sigma[A]$ (Definition 2.4), the Ruzsa distance $d(A, B)$ (Definition 2.5), and the additive energy $E(A, B)$ (Definition 2.8). We also introduce the concept of a partial sum set $A \overset{G}{+} B$ in Definition 2.28. The Fourier transform and the averaging notation $\mathbf{E}_{x \in Z} f(x)$, $\mathbf{P}_Z A$ is defined in Section 4.1, Fourier bias $\|A\|_u$ is defined in Definition 4.12, Bohr sets $\mathrm{Bohr}(S, \rho)$ are defined in Definition 4.17, and $\Lambda(p)$ constants are defined in Definition 4.26. The important notion of a Freiman homomorphism is defined in Definition 5.21. The notation for group theory (e.g. $\mathrm{ord}(x)$ and $\langle x \rangle$) is summarized in Section 3.1, while the notation for finite fields is summarized in Section 9.4.