CAMBRIDGE TRACTS IN MATHEMATICS

122 Duality in Analytic Number Theory

P. D. T. A. Elliott
University of Colorado

Duality in Analytic Number Theory

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/9780521560887

First published 1997

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

Library of Congress Cataloguing in Publication data

Elliott, P. D. T. A. (Peter D. T. A.)
Duality in analytic number theory / P.D.T.A. Elliott.
p. cm. – (Cambridge tracts in mathematics ; 122)
Includes bibliographical references and indexes.
ISBN 0 521 56088 8 (hardcover)
1. Number theory. 2. Duality theory (Mathematics) I. Title. II. Series.
QA274.73.B48 1996
512′.73 – dc20 95-48214 CIP

ISBN-13 978-0-521-56088-7 hardback
ISBN-10 0-521-56088-8 hardback

Transferred to digital printing 2005

Table of Contents

Acknowledgements

This book was largely written in my free time. I thank the American Mathematical Society and Cambridge University Press for allowing me to adapt Chapters two and three of my AMS monograph *On the Correlation of Multiplicative and the Sum of Additive Arithmetic Functions*. Partial summer support on National Science Foundation contract DMS 9300551 is reflected in some two hundred and seventy exercises that elaborate the text.

I thank David Tranah, editor of the Cambridge University Press, for his challenging advice.

I particularly thank Liz Stimmel who typed the whole project with virtuoso handling of TeX and its distant relatives.

The yellow sea cycled into a grey swell, lifting bergs of printout, feet deep. For the third time Jean heard the cry of the Dutchman, saw him break through the mist ....

Preface

In this book I have two aims. My first is to give a coherent account of a general method in analytic number theory, and to develop that method sufficiently far that it solves problems otherwise beyond reach. The method applies the simplest notions from functional analysis, and has its roots in geometry.

My second aim, bound to the first, and to me of equal interest, is a light discussion of the creation of the method as a raising of the underlying philosophical motivation into consciousness. In particular, this offers a paradigm for the application of the method itself.

I wrote the present work and my memoir: *The Correlation of Multiplicative and the Sum of Additive Arithmetic Functions* together. To facilitate a bridge between the two works I have elaborated the treatment of approximate functional equations given in Chapters 2 and 3 of the monograph. In particular, I preserve the same notation. For permission to do this I thank both the American Mathematical Society and Cambridge University Press.

The memoir applies the method to a problem not treated in this book. Background details in the construction of the method are omitted. Consideration of the problem to hand remains paramount. A large number of auxiliary results are required.

The present work is quite different in nature. The method itself is the object of study. Essential inequalities are derived in detail. In the body of the text background results either illustrate a mathematical idea, and can be omitted at first reading, or they are chosen to simplify the presentation of a proof, and can often be replaced by something weaker and more easily accessible. Otherwise only the basics of functional analysis on vector spaces of finite dimension, some elementary number theory, a little Fourier analysis and a familiarity with Cauchy's theorem in complex analysis are assumed.

Notation

A function is *arithmetic* if it is defined on the positive integers. Unless otherwise stated, arithmetic functions will be complex valued. The arithmetic function **1** is identically 1.

n, p generally denote a positive integer and a positive prime, respectively.

q often denotes a prime power, q_0 the prime of which it is a power.

$g(n)$ is *multiplicative* if $g(ab) = g(a)g(b)$ whenever (a, b), the highest common factor of a and b, is 1.

$f(n)$ is *additive* if $f(ab) = f(a) + f(b)$ whenever $(a, b) = 1$. A strongly additive function satisfies $f(p^m) = f(p)$, $m = 1, 2, \ldots$.

Additive and multiplicative functions are determined by their values on the prime powers. For completely multiplicative or additive functions we may suppress the condition of coprimality.

I employ the following standard arithmetic functions, the first three of which are multiplicative:

$\phi(n)$ of Euler, the order of the group of reduced residue classes $(\mathrm{mod}\, n)$,

$\tau(n)$, Dirichlet's function, the number of positive divisors of n,

$\mu(n)$ of Möbius, one at 1, otherwise zero when n has a squared factor, $(-1)^k$ when n is the product of k distinct primes,

$\Lambda(n)$ of von Mangoldt, $\log p$ when $n = p^m$, $m = 1, 2, \ldots$, zero otherwise,

$\pi(x)$, the number of primes not exceeding x.

An initial account of these and related functions may be found in Hardy and Wright [96].

In the Introduction, Chapters 19 and 34, $\tau(n)$ will also denote Ramanujan's modular coefficient function.

$\pi(x, D, \ell)$ denotes the number of primes, not exceeding x, which lie in the residue class $\ell (\mathrm{mod}\, D)$.

$s = \sigma + i\tau$, $\sigma = \mathrm{Re}\,(s)$, is a complex variable.

$L(s,\chi)$ denotes the standard Dirichlet series formed with the Dirichlet character χ, $\zeta(s)$ the Riemann zeta function. Classical results concerning these and related functions may be found in Davenport [19], Prachar [134].

$f(x) = O(g(x))$, $f(x) \ll g(x)$ both denote that $|f(x)| \le Ag(x)$, for some constant A, holds uniformly on a specified set of x-values. When qualified 'as $x \to \infty$', the set is a half-line $[x_0, \infty)$ with an undisclosed x_0.

$O(g(x))$ denotes a function f that satisfies $f(x) \ll g(x)$.

$f(x) = o(g(x))$ as $x \to \infty$, means that $\lim_{x\to\infty} f(x)g(x)^{-1} = 0$.

$o(g(x))$ denotes a function f that satisfies $f(x) = o(g(x))$.

Various natural generalisations of these notions are self-evident.

I shall many times employ the elementary bounds

$$\sum_{n \le x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right), \quad x \ge 1,$$

with Euler's constant γ, and

$$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1), \quad \sum_{p \le x} \frac{1}{p} = \log\log x + c + O\left(\frac{1}{\log x}\right), \quad x \ge 2,$$

where c is constant. These estimates are largely established in Hardy and Wright [96], Chapter XXII, in part employing the bound $\pi(x) \ll x/\log x$, $x \ge 2$, of Chebyshev. See, also, Chapter 15 of the present work.

$\nu_x(n; \cdots)$ denotes the frequency $[x]^{-1}N$, where $[x]$ counts the positive integers up to x, N the number of such integers n for which property $\cdots$ holds.

On the space of functions $f : S \to \mathbb{C}$, measurable with respect to μ,

$$\|f\|_\alpha = \left(\int_S |f|^\alpha d\mu\right)^{1/\alpha}, \quad \alpha \ge 1,$$

as usual. $1/\alpha' = 1 - 1/\alpha$, suitably interpreted when $\alpha = 1$.

Families of basic spaces, together with their duals, are introduced as follows:

$\mathbb{C}^s$, $L^\alpha(\mathbb{C}^s)$, $L^\alpha \cap L^\beta(\mathbb{C}^s)$, $\mathbb{C}^t$, $M^\alpha(\mathbb{C}^t)$, in Chapter 3,

$J^\alpha(\mathbb{C}^u)$, $J^\alpha \cap J^\beta(\mathbb{C}^u)$, $K^\alpha(\mathbb{C}^v)$, in Chapter 25, pp. 211–212,

$\lambda^\alpha(F)$, $\lambda^\alpha \cap \lambda^\beta(F)$, $\mu^\alpha(G)$, in Chapter 25, pp. 222-223.

E denotes expectation in the sense of probability theory in Chapters 25, 26, 28.

E denotes a shift operator in the Introduction and Chapters 23, 30, 31; and (with a slightly changed definition) again a shift operator in Chapters 32, 33.

In a composition TS of mappings

$$\xrightarrow{S} B \xrightarrow{T}$$

it is assumed only that the range of S is contained in the domain of definition B of T.

Introduction

Du Doppeltgänger! du bleicher Geselle! H. Heine.

In this volume I give a unified account of a method in the analytic theory of numbers: *the method of the stable dual.* The method is particularly effective in the study of arithmetic functions possessing algebraic structure.

The reader may check details of selected applications of the method; turn to the continuing developments discussed in the penultimate chapter; press on to the new. However, works in analysis, especially in analytic number theory, can seem formless. The leading thread, as Hadamard would have called it, [91] p. 105, becomes obscured by a mass of detail. All too often the conclusion will appear atop a pyramid of small steps, each step apparently insignificant. Sometimes this is due to the nature of the subject; sometimes it is not.

As footnote 2 on page 136 of his book [114], Lakatos makes the following remark:

> *Rationalists doubt that there are methodological discoveries at all. They think that method is unchanging, eternal. Indeed methodological discovers are very badly treated. Before their method is accepted, it is treated like a cranky theory; after it is treated as a trivial commonplace.*

Experience has convinced me of the validity of this statement. Analytic number theory is a dynamic subject, with partial results as moments of clarity. There rarely comes a static final result. Indeed, analytic number theory is largely concerned with method. An heterogeneous nature, allowing opportunities to exhibit frailties of human psychology, gives it a forbidding aspect.

Accordingly, from time to time I discuss not only the details of a proof, but also an intent of its argument, or the manifestation of a background philosophy. I set these comments in historical context, account the occasional false step, and indicate unsolved problems. In short I give not only an account of these results, but also of the construction of their proofs. I believe that most readers will want not only to understand, but also to create.

The inequality of the Large Sieve has its origins in a paper of Linnik, [120]. He applied Fourier analysis in a manner derived from the Hardy–Littlewood circle method. Subsequent papers: Rényi [142], Roth [145], Bombieri [6], Davenport and Halberstam [20], Gallagher [89], are important references, and the list is by no means complete. In this introduction I am concerned only with the fact that all of these treatments were Fourier analytic. Under this umbrella I include Rényi's use of the theory of probability.

In the Fall of 1962 I entered Trinity College, Cambridge, as a graduate student of H. Davenport. For my first semester I was supervised by A. E. Ingham. Almost at once I became interested in the inequality of the Large Sieve. Although I had already noticed it, Ingham brought to my attention Rényi's paper on the representation of an even integer as the sum of a prime and an almost prime, [141]. There Rényi applies a version of the Large Sieve together with Fourier analysis on the complex plane in the manner of classical analytic number theory. So matters stood.

Things intervened. My thesis for the doctorate in large part concerned integral zeros of cubic forms. However, I continued to apply the inequality of the Large Sieve, and began again to ponder its nature. I obtained a connection between the Large Sieve and the Turán–Kubilius inequality, [27], and corresponded with Rényi himself.

During the academic year, Fall 1969–Summer 1970, I was a visitor to the University of Colorado, Boulder, and as a visitor gave a Colloquium in January of 1970, on the Large Sieve. I summarised my interpretation of the Large Sieve at that time: *Inequalities of Large Sieve type come in pairs, the inequality and its dual* (*or conjugate*); *to establish such an inequality is to determine the spectral radius of a self-adjoint operator.* An account appears in [30].

Once seen in functional analytic terms, as an application of the notion of duality, the inequality of the Large Sieve becomes redolent of the general scheme of analytic number theory otherwise begun by Euler and developed by Dirichlet: Study the integers through Fourier analysis on an appropriate group.

I was more interested in the notion of duality itself, and the precise logical rôle that it played or might play in the proofs of arithmetical propositions. Familiar with projective geometry, where the duality is between point and line, I knew that Desargues' theorem, which is of course an axiom, has its converse as dual. Since application of inequalities of Turán–Kubilius form developed information on the integers from information on the primes, so the dual of the Turán–Kubilius inequality, directly related to the Large Sieve, might yield properties of the primes from those of the integers. Mathematically expressed, the directions *primes* $\rightarrow$ *integers*, *integers* $\rightarrow$ *primes* were dual. Moreover, notions and arguments might be similarly paired. For example: *operator corresponds to sufficiency, dual operator corresponds to necessity.*

Readers familiar with functional analysis will have observed that the second dual of an operator is often essentially the operator itself. According to the scheme of the example, proposition and dual proposition would then be equivalent; the combined result would have a finished form. This is indeed manifest in certain applications of the method of the stable dual. A proposition concerning an arithmetic function on the integers is valid if and only if another proposition concerning that function holds on the prime powers. Moreover, successive applications of the background philosophy allow loosely formulated initial propositions to be made increasingly precise, as well as suggesting their proof. The argument moves in tightening loops.

In this manifestation the method of the stable dual recalls *Analysis* from the mathematical thought of ancient Greece, a method briefly summarised by Euclid in his *Data*. At home in geometry, that method would here be augmented by the systematic application of the notion of duality, and realised in functional analytic terms. But is an operator not a tangent?

By the seventeenth century the Greek method of Analysis had been abandoned; it was insufficiently effectual. The quest for certainty came to prevail over the quest for finality; Lakatos, [114] footnote to p. 64. Besides this, unswerving fealty to the notion of finality also allows interesting and non-trivial partial results to go unappreciated. Even in projective geometry, my experience is that without further argument the application of duality in pursuit of a converse proposition may not suffice. The method of the stable dual can be enhanced by employing results derived under other aesthetics. Indeed, its form suggests this.

Chapters 1 to 3, 5 and 6 of the present work contain an initial account of the method of the stable dual. Chapter 3 introduces the Banach spaces $L^\alpha(\mathbb{C}^s)$, $M^\alpha(\mathbb{C}^t)$, $\alpha > 1$ defined on intervals of the prime powers, and the integers, respectively. Beginning with these spaces I construct others on which the Turán–Kubilius inequality and various generalisations of it are interpreted as bounds for operator norms. The formulation of the inequalities and the construction of the spaces go hand in hand.

The various spaces depend upon at least one real valued parameter. Moving this parameter I introduce the notion of stability.

Applications of the method follow.

Spaces $\mathcal{L}^\alpha$ of arithmetic functions are introduced in Chapter 8. These spaces embrace many functions of active interest in number theory.

In Chapter 8 I characterise the additive functions which belong to a given space $\mathcal{L}^\alpha$.

In Chapters 9, 10 and 11 the multiplicative functions which lie in $\mathcal{L}^\alpha$ and possess a non-zero (limiting) mean value are characterised. To formulate an appropriate generalisation of the classical result of Delange, [21], is part of the problem.

The results of these four chapters are in a sense final; equivalent propositions are obtained.

Chapter 13 contains short proofs of theorems of Wirsing, [170], and Halász, [92], concerning multiplicative functions with values in the real interval $[-1, 1]$, and in the complex unit disc, respectively, and having asymptotic mean value zero.

The second of these theorems is applied in Chapter 16 to obtain Erdős' characterisation of finitely distributed additive functions. The proof uses Fourier analysis.

Chapter 17 employs the method of the stable dual together with the results of Chapter 16 to characterise those multiplicative functions which belong to $\mathcal{L}^\alpha$ and do not have mean value zero. The theorems of this chapter are also largely final. They may be applied to give necessary and sufficient conditions for a multiplicative function in $\mathcal{L}^\alpha$ to have (asymptotic) mean value zero. Aesthetically this result is not quite as satisfactory as the theorem obtained for non-zero values. In studying the value distribution of an arithmetic function the fact that it has asymptotic mean value zero is not particularly helpful. The function might be identically zero.

Chapter 19 is an anecdote which applies results of Chapter 17 to Ramanujan's modular function coefficient $\tau(n)$.

The space $L^2(\mathbb{C}^s)$ introduced in Chapter 3 can be given an inner product; it becomes a Hilbert space. The operator A_2, there attached to the classical Turán–Kubilius inequality, has an adjoint, A_2^*. The spectrum of the self-adjoint operator $I(\text{dentity}) - A_2^* A_2$ can be largely determined. As I show in Chapters 20 and 21, this allows a localised version of the theorem of Chapter 8 to be obtained that is in a sense final. Underlying the classical Turán–Kubilius inequality and certain of its generalisations are approximate isometries. Within the aesthetic of probability, the particular case $\alpha = 2$ was first obtained by Ruzsa, [150], using complex Fourier analysis.

The characterisation of additive functions f for which the difference (discrete derivative) $f(n) - f(n-1)$ belongs to $\mathcal{L}^\alpha$, with its implicit connection to the correlation of multiplicative functions, apparently lies deeper than that for a plain additive function. Motives for such an aim together with ideas for its attainment are given in Chapter 23. The application of an inequality of Large Sieve type is suggested, and a short discussion given to exemplify possible procedures.

The case that $f(n) - f(n-1)$ belongs to $\mathcal{L}^\alpha$ for some α in the range $1 < \alpha \le 2$ is particularly interesting. In the notation of Chapter 3, underlying the study of additive functions in such $\mathcal{L}^\alpha$ is the chain of operators

$$(L^2 \cap L^{\alpha'}(\mathbb{C}^s))' \xrightarrow{A_\alpha} M^\alpha(\mathbb{C}^t) \simeq (M^{\alpha'}(\mathbb{C}^t))' \xrightarrow{A'_{\alpha'}} (L^2 \cap L^{\alpha'}(\mathbb{C}^s))'.$$

In the case $\alpha = 2$, $A_2' A_2$ may be replaced by the self-adjoint $A_2^* A_2$. In order to study the difference $f(n) - f(n-1)$ of additive functions in a similar manner it is necessary to interpose a difference operator between A_α and $A_{\alpha'}'$:

$$M^\alpha(\mathbb{C}^t) \overset{(I-E^{-1})}{\longrightarrow} M^\alpha(\mathbb{C}^t) \simeq (M^{\alpha'}(\mathbb{C}^t))'$$

where E^{-1} largely coincides with the classical forward shift of sequences. The operator $A_2^*(I - E^{-1})A_2$ is no longer self adjoint. Ideally, the effect of the operator $A_{\alpha'}' E^{-1} A_\alpha$ would be negligible. Although there is no initial reason for this to be true, an appropriate conjecture and evidence in its favour are given in Chapter 30. In the case $\alpha = 2$, the conjecture would be a particular yet more general inequality related to the theorem of Bombieri and Vinogradov, [6], and the conjecture of Elliott and Halberstam, [80], concerning primes in arithmetic progression. It would be an abstract norm inequality.

A partial validation of the conjecture, sufficient for many needs, is given in Chapters 27, 28. Satisfactorily small bounds are established for the norms of operators $PA_{\alpha'}' E^{-1} A_\alpha$, where the projections P preserve almost all of the space $(L^2 \cap L^{\alpha'}(\mathbb{C}^s))'$, so-to-speak. Additive functions are regarded as the convolution of two functions, one simple, the other supported on the prime powers. A Mellin transformation (in the complex plane) is used to strip off the simple function and an inequality of Large Sieve type applied to treat the function on the prime powers.

Fractional power Large Sieve inequalities appropriate for this purpose are developed in Chapter 25. It is difficult to see how these inequalities might be formulated or established without the notion of duality. Each desired inequality is the dual of an inequality involving high powers in mean, requiring in part a Riesz–Thorin interpolation. To facilitate a rapid proof I appeal to an inequality of Rosenthal, from the theory of probability. As its form suggests, this inequality can be readily established by adapting the methods of Chapter 2. I demonstrate so in a short series of exercises in Chapter 26.

In part, the treatment of differences of additive functions follows argument for a plain additive function, lifted onto larger spaces.

Chapter 31 contains analogues of the results in Chapters 20 and 21. The additive function $f(n)$ is replaced by its difference $f(n) - f(n-1)$. The inequalities obtained have the desired form but are not quite as sharply localised as their (earlier) models. They much improve an earlier result of Wirsing, [171]. Apparently the case $\alpha = \infty$ was conjectured by Ruzsa.

In Chapter 32 I employ the results of Chapter 31, together with a polynomial ring of shift operators, to characterise those additive functions f for which $\lambda_1 f(n + a_1) + \cdots + \lambda_k f(n + a_k)$ belongs to $\mathcal{L}^\alpha$, where $a_1, \ldots, a_k$

are distinct integers and the λ_j complex. This bears upon a conjecture of Kátai, [108]. The results of this chapter also have a final form; equivalent propositions are obtained. They may be compared with those of Chapter 8.

Apart from changes of detail, the results of Chapters 25, 27, 28, and 30 to 32, motivated by the ideas of Chapter 23, I had already established in 1987, [59], [63]. Owing to exigencies of publication, they first appear here.

In Chapter 34 I discuss a variety of theorems derived using the method of the stable dual, and offer exercises and comments related to parts of the text. It should perhaps be emphasised that whilst the results of Chapters 8 to 11, and 13 can also be obtained without application of the method of the stable dual, at present the general results of Chapters 31, 32 and many of the results in Chapter 34, cannot.

At appropriate locations in the text I have placed chapters that contain only exercises and comments. Not including Chapter 34, there are ten such chapters, of a more or less elaborate nature, together containing 250 exercises. At the expense of an occasional repetition I have rendered the main body of the text independent of these chapters. However, to omit them is to omit much. There I gloss problems and ideas considered in the text, discuss approaching difficulties and indicate connections with classical or modern technique in analytic number theory.

Exercises face variously from analysis towards number theory or from number theory to analysis; many are linked in a chain around a specific object. Thus according to Chapter 14, experience in the method of the stable dual allows a proof of Wirsing's theorem different from that given in Chapter 13, and which does not apply the prime number theorem. With this as catalyst, Chapter 15 deploys a series of exercises to telegraph a commentary upon the classical approaches to the prime number theorem.

Taken together with the main body of the text, the exercise chapters introduce a wide collection of notions that have been of advantage to me in an ongoing study of analytic number theory.

The present volume might be used as a text or for private study. In particular, I provide at once a chapter on the qualitative nature of duality and Fourier analysis. In a short course it would suffice to read section **5**, concerning the duality principle.