PREFACE

This book shows the scope of analytic number theory both in classical and modern directions. There are no division lines; in fact our intent is to demonstrate, particularly for newcomers, the fascinating countless interrelations. Of course, our picture of analytic number theory is by no means complete, but we tried to frame the material into a portrait of a reasonable size, yet providing a self-contained presentation.

We were writing this book in a period of time during and after teaching courses and working with graduate students in Rutgers University, Bordeaux University and Courant Institute. We thank these institutions for providing conditions for both of us to work together. We shared ideas on what this book should be about with many of our colleagues, who gave us critical suggestions. Among them we would like to mention Étienne Fouvry, John Friedlander, Philippe Michel and Peter Sarnak. During a long process of typing and preparation of this book for publication, we received stimulating encouragement and technical advice from Sergei Gelfand, for all of his help we express our gratitude. Carol Hamer helped to polish some of our English phrases while her little boys tried to destroy the TeX files without success. We thank them all for the output.

Henryk Iwaniec
Emmanuel Kowalski
15 December, 2003

INTRODUCTION

Analytic Number Theory distinguishes itself by the variety of tools it employs to establish results, many of which belong to the main streams of arithmetic. It is not part of analysis nor of any particular discipline of mathematics, however it does interact indeed with various fields. Therefore everybody seems to view the subject differently. This vast diversity of concepts of analytic number theory is its great attraction. Our desire in this book is to exhibit the wealth and prospects of the theory, its charming theorems and powerful techniques. However it is not our primary objective to give proofs of the strongest results, although in many cases we come quite close to the best possible ones. Rather we favor a reasonable balance between clarity, completeness and generality. The book was conceived with graduate students in mind so the reader will often find that our emphasis is on reasoning throughout the arguments. Of course our presentation is subjective, and in retrospect may lose its meaning. Certainly we do not always follow the original lines of discovery, but occasionally we do draw brief historical perspectives.

Leonard Euler must get credit for the first use of analytical arguments for the purpose of studying properties of integers, specifically by constructing generating power series. Euler's proof of the infinity of prime numbers makes use of the divergence of the zeta function and the corresponding product over primes, which is named after him. This was the beginning of analytic number theory. Next came P. G. L. Dirichlet whose creation of the theory of L-functions for characters, resulting in the proof of the infinity of primes in arithmetic progressions, makes him the true father of analytic number theory. From these early days to modern times the distribution of prime numbers constitutes the core of the subject. This will be apparent in the course of our book. The first two chapters cover questions of primes up to the elementary methods of P. Tchebychev.

Chapter 3 provides definitions and basic properties of Dirichlet characters and the Gauss sums. Characters on ideals of imaginary quadratic fields are also introduced, not only because they play a supporting role in subsequent chapters but to show a bit of analytic number theory beyond the traditional domain of the rational integers as well; there will be other examples throughout the book, for instance, elliptic curves.

Poisson summation for number theory is what a car is for people in modern communities – it transports things to other places and it takes you back home when applied next time – one cannot live without it. Chapter 4 presents a classical account of this basic technology. Many readers do realize now, others will figure out later, that we are already talking about ideas of modular forms. But we continue our considerations along traditional lines (both classical and more recent ones) before the concept of modularity takes the leading position.

The celebrated memoir of B. Riemann on the zeta function is embedded in the context of abstract L-functions in Chapter 5. It is not our style to consider things in terms more general than necessary, so defining a class of L-functions which suits minimum requirements of our forthcoming applications was not without difficulty and hesitation. In this way we could convey to dedicated researchers that generalizations are not always straightforward. For instance, to establish the zero-free region for L-functions of degree > 1 one cannot rely on the same principles as for the Dirichlet L-functions. The key ingredient is the Rankin-Selberg convolution. On the other hand, the problem of exceptional zero is resolved for many automorphic L-functions of degree > 1 (not without clever constructions) while it remains open for the L-functions with real characters. Furthermore in Chapter 5 a message is sent that a better life exists in the world of automorphic forms than in the zoo of degree one L-functions.

Analytic number theory does not mean non-elementary. The first author recalls that his first serious encounter with analytic number theory started by reading the lovely book of A.O. Gelfond and Yu.V. Linnik, "Elementary Methods of Analytic Number Theory". When an ambitious beginner starts from there her/his love of the subject is sealed forever. Try it yourself! One is instantly captured by sieve methods. In this book we do not have space to give justice to this marvelous idea, nevertheless Chapter 6 should suffice for basic applications.

Next comes the "Large Sieve", which is not a sieve but a name for other things. Yes, it did originate from a short paper by Linnik on a sieve problem, but it took time to recognize the true nature of these ideas. In Chapter 7 we reveal our viewpoint and the crucial attributes (spectral completeness, orthogonality), then we demonstrate the amazing power of the large sieve on selected old and new problems. Other features of the large sieve are scrutinized showing the good and the bad sides. For example, the approach using the duality principle is fruitful for harmonics of degree one (characters) while producing poor results for harmonics of larger degree (like for example the eigenvalues of Hecke operators). The controversy over the proper place of the large sieve is academic. Simply speaking the large sieve inequalities are parts of bilinear forms theory.

Estimates for exponential sums are the first tools which deeply penetrate the problems of analytic number theory beyond natural structures. These cannot be grasped by harmonic analysis alone. See what clever use people made of the property that a shifted interval is another interval, that adding an integer to a set of integers yields again a set of integers (sorry, primes are not preserved!). We challenge algebraists to find a structural explanation of the power of such arguments! They should read Chapter 8 to find what H. Weyl built out of these observations. Van der Corput and Vinogradov are also the main figures from the early stages of that discipline. A lot of work and talking went into our presentation of Vinogradov's method, because it is not quite correctly explained in numerous publications. At some point Vinogradov departs from the Weyl differencing process and treats multi-dimensional exponential sums as bilinear forms (this is the way we think of it anyway).

The next two chapters show more recent technology which was developed to replace the unproven Riemann hypothesis in applications to the distribution of prime numbers. We are talking about estimates for the number of zeros of L-functions in vertical strips which are positively distanced from the critical line. Hopefully in a future one will say we were wasting time on studying the empty set. Great ideas

are camouflaged there in arguments of enormous complexity, so this might not be enjoyable for everyone at first. However if you think the Riemann hypothesis is not provable in your lifetime, please read and admire these unconditional substitutes. Special mention goes to Hugh Montgomery, Martin Huxley and Matti Jutila for the most original contributions.

Although we are primary interested in rational integers one can learn and benefit a lot from arithmetic of other fields. Not only from the number fields or p-adic fields, but indirectly from the fields of finite characteristic as well. Particularly fruitful are the methods of exponential sums over the finite fields. In Chapter 11 we prove (among other things) the Riemann hypothesis for special curves which yields the celebrated estimate of Weil for Kloosterman sums. The Kloosterman sums have been employed to solve various problems of analytic number theory from the beginning of their creation in 1926. We also mention briefly the state of the knowledge of exponential and character sums over algebraic varieties. Applications of these are harder to make, yet there is a handful of examples in the literature. It was a painful decision to exclude all but the simplest from presentation in this book. Otherwise to do full justice for these highly sophisticated ideas we would have to choose the most complicated application for which we have no room. It suffices to say that a preparation of a given problem of analytic number theory to an estimate for character sum over varieties can be the state-of-the-art in its own right, never mind that the final argument is powered by the outside forces of algebraic geometry.

Dirichlet characters are already discussed in Chapter 3 and we return to them in Chapter 12 to treat very short character sums. Again one must be inventive to break limits of natural structures. Burgess theorem is a fine example.

Sums over primes are treated in the next chapter. When Vinogradov succeeded in estimating sums over primes of additive characters, which he needed for a solution to the ternary Goldbach problem in conjunction with the circle method, it was a shocking result. Before him the Grand Riemann Hypothesis could do the job, but keep in mind that the Riemann hypothesis is still not established. The original ideas of Vinogradov were borrowed from combinatorial sieve and were rather complicated. Recently developed identities offer much simpler treatments of more general sums over primes. As they share the same fundamental principle (reducing the sum to bilinear forms) the results are pretty much the same, so the choice of the method is a matter of taste and technical convenience. To capture the key elements in Chapter 13 we develop more than one identity.

A popular criterion for analytic number theory is that complex variable analysis is being used. Perhaps it is better to say harmonic analysis, since the action of the latter is more profound. For a long time, analytic number theory flourished exclusively from abelian harmonic analysis, that is to say from the Fourier transform in $\mathbb{R}^n$. There is still a great potential in this classical analysis to be explored. However much stronger fertilizers began to act on the soil of analytic number theory in recent times. These are automorphic functions. Of course, modular forms have been driving algebraic aspects of number theory much longer, but in a limited scope (confined to holomorphic forms). New resources of automorphic theory are found in the spectral analysis, the foundation of which was led by H. Maass and A. Selberg at the turn of the $1940's$ (real-analytic cusp forms, Eisenstein series, trace formula). In simple terms a non-abelian harmonic analysis found its role in analytic number theory. Truly effective expansion of spectral methods into analytic number theory began about twenty five years ago, changing the face of either subject irrevocably.

This book barely addresses the fascinating issues of the new direction throughout Chapters 14, 15, and 16. Our featured application is to estimation for sums of Kloosterman sums. This is a good choice (if no more can be accommodated), because the reader can appreciate the new tools by comparing with the earlier results derived in Chapter 11 by algebraic considerations. Another application of the spectral theory of automorphic forms to an arithmetical question is presented in Chapter 21, that is to the equidistribution of roots of quadratic congruences of prime moduli. The spectral theory continues to grow extensively, so it would be premature to wrap it up here or in any other book. For further reading we recommend [**I3**], [**Sa3**].

Although the spectral methods of automorphic forms predominate current research in analytic number theory, the traditional problems continue getting our attention with respectful intensity throughout the remaining chapters. Great treasures of the subject mustn't be buried in the past. First of all a newcomer should learn the stories of primes in arithmetic progressions to large moduli. In Chapter 17 she/he will find how E. Bombieri and A.I. Vinogradov bypassed the Riemann hypothesis to establish (by the large sieve and other means) unconditional results with applications as good as one can get from the RH itself. Of course our arguments are not identical with the original ones (of 1965) since we take advantage of later simplification, in great measure due to P.X. Gallagher.

Chapter 18 goes further back to 1944 when Linnik gave an extraordinary bound for the least prime in an arithmetic progression. For a long time this bound was considered as the most difficult theorem in analytic number theory. And yes, it is still hard by today's standards, and one can still learn a lot from the technology applied! Faced with the obstacle of the exceptional zero, Linnik brings the repulsion effect (he calls it Deuring-Heilbronn phenomenon) to a new level; amazingly enough he turns the problem to his advantage! This is a fascinating development in the history of analytic number theory which we recommend one should master for a better understanding of the status of the exceptional zero today.

Once upon a time the famous Goldbach problem was worth a million dollar prize award. For applications the problem (representations of even integers by the sums of two primes) has no great merit, but as an intellectual challenge one would be proud to crack it. Probably something new about prime numbers would be revealed then. Read Chapter 19 to improve your chances.

Chapter 20 is serious. Here analytic methods storm the domain of diophantine equations, which from ancient Greeks was exclusively a business of arithmetic. Started by Hardy - Ramanujan, continued by Hardy - Littlewood and developed substantially further by Kloosterman, the circle method uses orthogonality of additive characters to detect equations, not only to solve algebraic equations but a large class of additive problems over special integers as well. The toughest are the binary additive problems. They are not completely solved by the Kloosterman method, but at least we get a very reliable picture of what the true asymptotic for the number of solutions should be. Kloosterman sums which we covered in the preceding chapters are instrumental in the circle methods. After classical ideas we propose a more direct variant which in principle should produce the same results, however without employing Kloosterman sums. One should read Chapter 20 with an open mind, separate technical (still attractive) elements from conceptual devices to see clearly

the connections with modular forms. Certainly Kloosterman and Rademacher were aware of these intrinsic connections, while they are overlooked by some specialists in the circle method.

Equidistribution problems for sequences of special integers, lattice points in various domains, solution to diophantine equations, etc, constitute a heavy industry over the analytic number theory. We regret there is no space to run this industry in full capacity in the book. The book of M.N. Huxley [**Hu4**] treats only the lattice point problems, however quite deeply. In Chapter 21 we are dealing with the problem of distribution of roots of a quadratic equation reduced modulo prime. As the prime modulus tends to infinity we show that the roots are uniformly distributed. The arguments include almost everything that we developed in the book so far, thus showing that the industry is robust.

Because of failure of the unique factorization of algebraic integers, the arithmetic of number fields is not as easy as for rational numbers and sometimes perplexing. The complexity is measured by the order of the ideal class group. Naturally the case of imaginary quadratic fields received the first and the most attention because units do not interfere. We do know that the class number grows to infinity (so there is only a finite number of imaginary quadratic fields with a fixed class number), but the serious issue is to estimate the class number effectively. Chapter 22 describes the problem thoroughly and prepares the ground for the advances in Chapter 23. The effective lower bound for the class number (due to D. Goldfeld) may not appear strong for demanding researchers, yet it is deep with respect to results taken from other sources. First of all it uses the Gross-Zagier formula for L-functions of elliptic curves at the central point. We do provide a substantial overview of the involved arguments from elliptic curves, although these are more geometric than analytic. The analytic arguments themselves are quite delicate. Actually they came first, the L-functions of elliptic curves being supplementary. Indeed we worked out an effective lower bound for the class number which depends on the order of vanishing of general L-functions of degree two, suspecting that the requirements are satisfied by quite a few of them.

In Chapter 24 we prove a very classical result of Selberg that a positive proportion of zeros of the Riemann zeta function lies on the critical line. This is a good place to learn about the mollification techniques (a kind of smoothing), which is used in many works today and will reappear in Chapter 26.

Assuming the Riemann hypothesis, H.L. Montgomery revealed in 1974 that the distribution of zeros of $\zeta(s)$ follows the behavior of eigenvalues of certain "ensembles" of unitary matrices. More recently physicists joined the team of workers in number theory, creating a new excitement and hope for finding a path to a proof of the Riemann hypothesis. This is the main objective of the so-called random matrix theory, one of the most popular subject and driving forces of current analytic number theory. It offers reliable models for predicting the behavior of arithmetical quantities which for a long time were shrouded in mystery. The consistency of the random matrix theory with the harmony of integers still seems quite surprising. Whatever the future of this enterprise will be, due to the current cooperation, analysis is closer to arithmetic than ever before. A subject of such magnitude cannot be fully presented in a short space. Therefore in Chapter 25 we stick to the original theme of the correlation of zeros of $\zeta(s)$ and its variations on the zeros of

families of automorphic L-functions which are near the central point. We leave it for the reader to judge whether the ideas of random matrix theory are realistic for launching an attack on the Riemann Hypothesis.

In recent investigations the central values of L-functions appear in a variety of formulas with vanishing or non-vanishing assumptions. Take for example [**IS2**] where an effective lower bound for the class number of imaginary quadratic fields is derived essentially from the non-vanishing of central values of families of L-functions, to the contrary of the vanishing requirements in the previous investigations. Another example is the formula of T. Watson [**Wa**] by means of which the quantum-ergodicity conjecture (that is the equidistribution of Maass cusp forms) is reduced to a subconvexity bound for certain L-functions of degree four. We consider in detail one non-vanishing statement which has applications to arithmetic geometry.

We hope this book will show the picture of analytic number theory in plenty of colors. However we must say that a lot of significant topics are left out. Missing are the dispersion method, the amplification method (see [**M2**]), some analytic techniques from diophantine approximations and transcendence. Moreover probability arguments are barely exposed, and we didn't touch ergodic theory either, whose impact on number theory has been felt strongly in the last years.

We also try to show some details of the powerful theories which are developing as the most useful new tools for analytic number theory, in particular the theory of higher-degree automorphic forms and their L-functions, and algebraic geometry; young researchers in particular should be encouraged to develop expertise in these subjects. It is certain that spectacular applications have only begun and more will be open to those who understand both sides. Dually, arithmetic geometry and algebraic number theory also give and promise a wealth of new questions, or new aspects of old ones, where the skills and techniques of analytic number theory will be tested to the utmost. Hopefully they will bring rich rewards to those who will try to come to these open fields... We barely mention some questions related to elliptic curves but we believe that there is much more to discover. The deep conjectures of Lang and Trotter [**LT**] are already quite popular, and a few other challenging problems may be found in [**Ko1**].

The exercises inside each section serve a dual purpose, some are to improve the reader's skill, the others serve as additional information about the subject. Historical remarks are brief, to give some orientation in the development of the matter, rather than to credit exhaustively the inventors. The only advice we offer to new researchers is read! read! read! many papers with complete proofs. Knowing a result in analytic number theory is only the first step to liking it; more important and rewarding is to understand the arguments of its proof. Our viewpoint is that making mathematics should not be rated like breaking sport records. Sometimes the strongest result is boring while a slightly weaker one generates great pleasure.

Formal prerequisites for much of the book are rather slight, not going beyond differential calculus, complex analysis and integration, especially Fourier series and integrals. It is more important for the reader to have or acquire a good understanding of how to manipulate inequalities and not simple identities.

In later chapters automorphic forms become important. We have included two survey chapters, yet we expect that many readers will have already some knowledge of this important topic, or will study it independently.

In some sections (for instance Sections 5.13 and 5.14), which are intended as convenient references for certain facts and results which are hard to locate in proper form in the literature, we assume that the reader has some familiarity with other topics, such as representations of groups and algebraic geometry.

Sections of this book were written over a period of time, therefore readers will notice a slight change of style and repetition. We think that a small redundancy is helpful for reading long arguments. Occasionally the same object is introduced again in a different chapter in local terminology which should be more familiar in a particular context. We believe this flexibility is justified for comfort, even at the expense of losing uniqueness.

Our notations are mostly standard. But since inequalities with unspecified constants are the lifeblood of analytic number theory, and since there are sometimes controversies on this subject, we spell out the meaning of the various comparison symbols $O()$, $o()$, $\sim$, $\asymp$ or $\ll$. Most important, we use Landau's $f = O(g)$ and Vinogradov's $f \ll g$ as synonyms; thus $f(x) \ll g(x)$ for $x \in X$ (where X must be specified either explicitly or implicitly) means that $|f(x)| \leqslant Cg(x)$ for all $x \in X$ and some constant $C \geqslant 0$. Any value of C for which this holds is called an implied constant. Since a constant is most often a function looking for a variable, the "implied constant" will sometimes depend on other parameters, which we explicitly mention at the most important points (but sometimes it is clear from context). If there is no other dependency, we speak of "absolute constants". This usage means that our $O()$, for instance, is not the same as that of Landau or Bourbaki. We use $f \asymp g$ to mean that both relations $f \ll g$ and $g \ll f$ hold, of course with possibly different implied constants.

However $f = o(g)$ for $x \to x_0$ means that for any $\varepsilon > 0$ there exists some (unspecified) neighborhood U_ε of x_0 such that $|f(x)| \leqslant \varepsilon g(x)$ for $x \in U_\varepsilon$. Then $h \sim g$ means $h = g + o(g)$. Those are the same as in Landau or Bourbaki.

Among the few notation which may be unfamiliar to a beginner, we mention that $p^k \| m$, where p is prime and k an integer, means that p^k divides m exactly (i.e. p^{k+1} does not divide m). The integral part $[x]$ is the integer n such that $n \leqslant x < n + 1$.

We sometimes use the notation $\sum^\star$ to denote sums restricted to a subset of "primitive" objects, which will be indicated in each case, and $\sum^\flat$ to denote a sum restricted to squarefree numbers.