ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 15, Number 2, Spring 1985

APPLICATIONS OF SPECTRAL THEORY TO NUMBER THEORY

JAMES LEE HAFNER

In Memory of E. G. Straus and Robert Smith

The essence of this talk is fundamentally philosophical. That is, what I would like to do is describe for you in some general terms why spectral theory should have anything to do with number theory. I am directing this primarily to those number theorists who would like some general picture of how some of these various theories fit together; this is to provide a philosophical basis for a beginning study of these subjects. The point I will try to make is that there is a need and desire when solving problems to appeal to the larger contexts in which these problems may exist, those that go beyond the language in which the problems are stated. Analogously, one should keep in mind the power of complex variables in the theory of primes.

I would like to emphasize that the philosophy contained herein is strictly my own and reflects only my limited mathematical perceptions and experience. There surely are aspects of the general picture which should be included and are not; these omissions should only indicate my general ignorance and not a prejudice or judgment against their value or position.

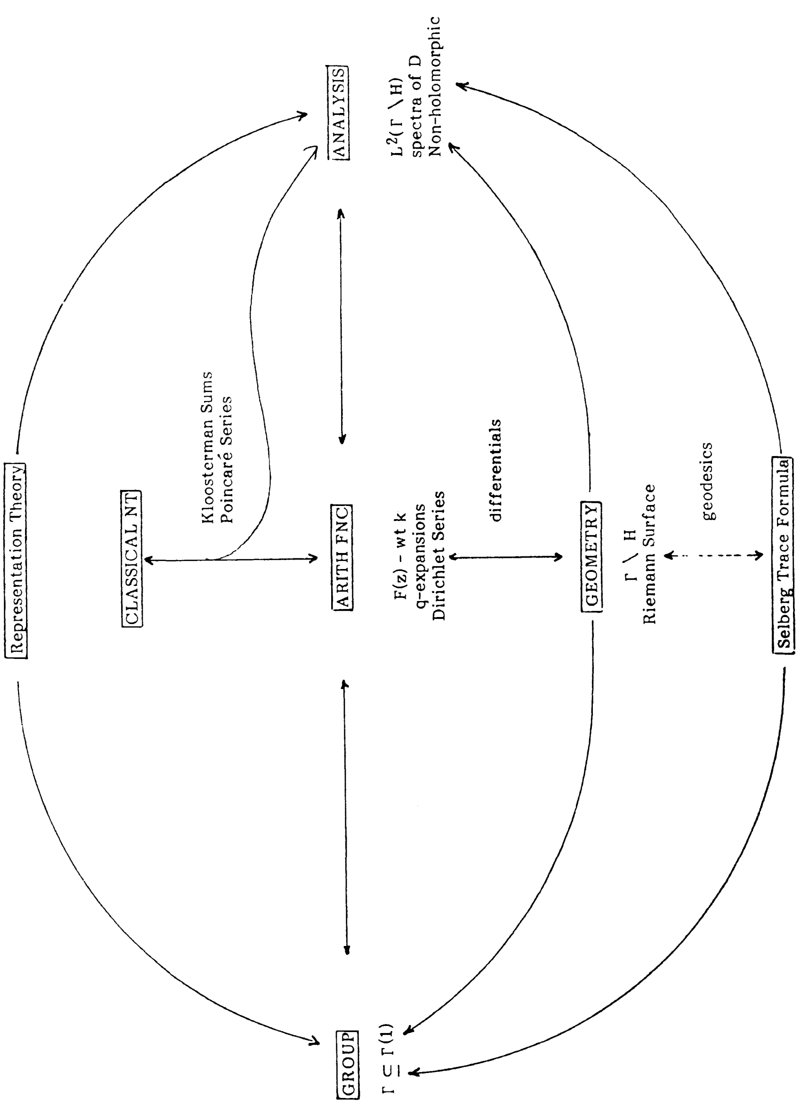
In the first part of this talk, I will present graphically some basic interrelationships between various subjects in number theory. In the second part I will sketch briefly some of the major theorems of recent years which are indicative of the power of appealing to these relationships. These results will be primarily due to others. I should indicate that the references given are examples and should not be taken as an exhaustive bibliography of these subjects.

With these preliminary remarks and apologies out of the way, let me proceed. The diagram on the next page indicates some relative connections between various subjects. Let me explain briefly what this diagram means. (The words in boxes indicate the general mathematical field involved.)

We begin on the left with a group Γ contained in PSL $(2, Z) = \Gamma(1)$,

Received by the editors on June 30, 1983.

the modular group. (We could, more generally, consider a Fuchsian group of the first kind (with parabolic elements) contained in PSL(2, R).) Here some of the basic questions one might ask involve the general structure of the group. In particular, a description of the conjugacy classes, especially the elliptic and parabolic classes might be desirable. This should also include a study of the quotient groups and their arithmetic. See Rankin [25, Ch. 1].



Associated to that horizontally in the diagram is the theory of (meromorphic) modular forms and functions for the group Γ, especially cusp forms of a fixed weight k. These are functions F which are meromorphic on the upper half-plane H and which transform according to the rule

$$F(Mz) = F\left(\frac{az+b}{cz+d}\right) = (cz+d)^k F(z)$$

for all $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in Γ. For example we have Ramanujan's Δ-function given by

$$\begin{aligned} \Delta(z) &= q \prod_{n=1}^{\infty} (1-q^n)^{24} \\ &= \sum_{n=1}^{\infty} \tau(n)\, q^n, \qquad q = \exp(2\pi i z) \end{aligned}$$

which has weight 12 and $J(z)$, the elliptic modular invariant of weight zero, both of whom live on $\Gamma(1)$. See Apostol **[1]**. The classical theta functions are also included here. In this part of the theory we study the coefficients of the so-called Fourier series or q-expansions and their associated Dirichlet series. It is here that we meet the Hecke correspondence described in Ogg **[23]**. The Hilbert space of cusp forms of fixed weight k and the theory of holomorphic Poincare series also arises in this context. Of course we should not ignore the theory of the Hecke operators which give one of the natural connections between the group theory on the left and the arithmetical functions in the middle. Furthermore, the group Γ on which these forms live reflects itself in the multiplicative behavior of the coefficients.

Next we move down simultaneously from these two positions to the area marked geometry. Here we consider Γ acting on the upper-half plane H by linear fractional transformations and we form the orbit space $\Gamma\backslash H$ (or a fundamental region with the appropriate sides identified). This space can be made into a Riemann surface with choices of local parameters which reflect the structure of the group. For example, at a cusp there is infinite ramification and at elliptic points the ramification equals the order of the elliptic point. (This is ramification when passing from H as a covering suface to the quotient $\Gamma\backslash H$.) As is described very thoroughly in Lehner **[22]**, Ch. VI], there is also a connection between differentials and meromorphic functions on the surface and the above Hilbert space of cusp forms. Via the Riemann-Roch Theorem, this can be exploited to give the dimension of these spaces.

Next we move to the right into the much more purely analytical topics of the Hilbert space of square integrable functions (automorphic functions) on $\Gamma\backslash H$. The beauty of this subject is that on the one hand it carries implicitly a large portion of the information of the previous two topics (with all weights treated uniformly) but also allows an appeal to the general theory of harmonic analysis. It is here that we finally discover spectral theory for the Laplace-Beltrami operator

$$D = y^2(\partial_x^2 + \partial_y^2)$$

and its various deep questions about the existence and exhibition of non-holomorphic (real-analytic) modular eigenforms (Γ-invariant eigenfunctions of D). This is really the spectral resolution of D which we seek explicitly. Of course, the Hecke operator theory also plays a role here and so questions about the (classical) Fourier coefficients of these forms come into play.

Providing a solid foundation for this part of the diagram is one of the most beautiful and fundamental theorems of the last thirty years. Of course I am referring to the Selberg Trace Formula [**29**]. It exhibits an explicit connection between the harmonic analysis (spectral questions) of D on $L^2(\Gamma\backslash H)$ and the group structure which we began with. Symbolically, the trace formula takes the form

$$\begin{aligned}\mathrm{Tr}\, K &= \sum h(\text{eigenvalues})\\ &= \sum g(\text{conjugacy classes}),\end{aligned}$$

where K is a certain type of integral operator on the L^2 space and g and h are functions derived from the kernel of K. See also Kubota [**19**] or Hejhal [**13**].

Note that there is also an arrow connecting the trace formula to the Riemann surface with the word geodesics. That is, lengths of geodesics on the surface are reflected in the norms of conjugacy classes exhibited in the trace formula. See Example 5 below for more on this connection.

We have yet to connect these theories to classical number theory; so far we have only dealt with that part of number theory involved with modular forms. The connection with classical number theory is via Kloosterman sums as indicated in the upper portion of the diagram. Kloosterman sums arise naturally in a number of contexts. The first, from our point of view, is in the theory of holomorphic Poincare series mentioned above. This was a fundamental discovery of Petersson (see Rankin [**25**, Ch. 5]) and is intimately connected to the theory of modular forms. They also arise naturally in the theory of non-holomorphic Poincare series as well. These sums can therefore be studied in the context of the Hilbert space $L^2(\Gamma\backslash H)$. This was pointed out by Selberg [**30**]. For example, he showed that the Dirichlet series

$$\sum_{c=1}^{\infty} S(m, n, c)\, c^{-s} \tag{1}$$

has singularities related to the spectrum of the operator D. So here we find another clear connection between spectral theory, group theory and modular forms.

One famous theorem which I could mention here is due to Selberg **[30]**. He uses Weil's famous estimate **[33]** for Kloosterman sums to deduce the $\lambda(1) \geqq 3/16$ inequality. More precisely, if Γ is a congruence group, then the smallest non-zero eigenvalue of D is no smaller than 3/16. The conjecture of Selberg is that $\lambda(1) \geqq 1/4$, whereas general theory only yields that $\lambda(1)$ is positive. See also Randol **[24]**. This has consequences to the analytic continuation of the Dirichlet series in (1).

These Kloosterman sums arise also in a whole range of classical number theoretic problems. The most naive way to see why this should be so is to recognize that Kloosterman sums are quite general exponential sums, or alternately they can be viewed as finite Fourier series. Thus, for example, any problem which involves Dirichlet characters might be treatable via Kloosterman sums. I'll list a few of the more important problems which can be so viewed.

—the greatest prime factor of $n^2 + 1$

—the number of divisors of $n^2 + 1$

(2) —power mean values of $\zeta(1/2 + it)$

—the Brun-Titchmarsh theorem

—the greatest prime factor of $p + a$.

Their connections to Kloosterman sums in many cases are due to Hooley and others. See the bibliography of Iwaniec **[18]** for some references.

Finally, I should put a roof over this diagram. Encompassing many of these interrelationships is the general representation theory. In the context of local theory for algebraic number theory (or algebraic varieties), the local and global L-series are, via the Weil correspondence, associated to holomorphic and non-holomorphic modular forms. An analysis of these L-series produces information about their coefficients which leads to a better understanding of the forms themselves. This part of the theory has been formulated and formalized by Jacquet, Langlands, Serre, and many others. I should mention the reference to Gel'fand, Graev and Piatetskii-Shapiro **[4]**.

Let's not forget the work of Deligne and Grothendick in algebraic geometry which lead to a proof of the Ramanujan-Petersson conjecture

$$|\tau(n)| \leqq d(n)\, n^{11/2}.$$

Let me now take the liberty to mention in rough detail some specific results which illustrate the connections described in the diagram.

EXAMPLE 1. We should first mention some further results and ideas of

Selberg [**30**]. In this paper, Selberg indicates a proof of the eigenvalue inequality mentioned above. He also outlines some general methods for dealing with sums of the type

$$\sum_n \tau(n)\,\tau(n+k) \text{ or } \sum_n d(n^2+1). \tag{3}$$

By generalizing the Rankin-Selberg convolution (replace the Eisenstein series by an appropriate non-holomorphic Poincare series) one can show that the Dirichlet series associated to these sums have singularities at points related to the eigenvalues of D and that the residues can be calculated. This is similar to the situation in (1). This provides a powerful and computational tool for analyzing these sums.

EXAMPLE 2. Good ([**5**] – [**9**]) has made remarkable use of these ideas. In [**5**] and [**7**] he uses some rather complicated expansions and some very deep estimates from the spectral theory to obtain a very good asymptotic expression for

$$\int_0^T |L_\tau(6+\mathrm{it})|^2\,dt$$

where

$$L_\tau(s) = \sum_{n=1}^{\infty} \tau(n)\,n^{-s}. \tag{4}$$

This leads to non-trivial estimates for $L_\tau(6+\mathrm{it})$, a problem analogous to estimating $\zeta(1/2+\mathrm{it})$. In [**6**] he deduces quite remarkable estimates for the Fourier coefficients of holomorphic modular forms for an arbitrary Fuchsian group of the first kind. These are similar to, though (as should be expected) not as effective as, Rankin's early results [**26**] for τ.

EXAMPLE 3. In a remarkable paper [**31**], Thompson connects group theory and the function theory of the Riemann surfaces $\Gamma\backslash H$ to prove that certain finite simple groups are Galois groups of algebraic number fields over the rationals (or at worst over a subfield of an explicit cyclotomic field). In his proof, he does a delicate balancing act between three (Fuchsian) groups $\Delta \subseteq \Lambda \subseteq \Gamma$ and the fields of meromorphic functions on the associated Riemann surfaces. He views the quotient groups as Galois groups for the function field extensions. These quotient groups have the property that they are direct products of copies of the finite simple group in question. By passing back and forth between the analytic and the algebraic ideas he can exploit the inherent structure of each to derive information about the other. Then, by what I consider brute force, he brings this down from function fields to algebraic number fields. As a special case he shows that the "friendly giant" Fischer-Greiss group is a Galois group over the rationals.

EXAMPLE 4. Here I only want to mention briefly the work of Iwaniec, Deshoullier, Fouvry and others (see **[18]**). In these papers, a significant amount of progress has been made in many of the classical problems mentioned above in (2). They proceed first by considering a reduction of the problem to sums of Kloosterman sums. The basic idea is to exploit the inherent cancellation present in such sums. Using ideas from the Kuznetsov-Bruggerman Trace Formula (**[2]** and **[20]**) which ties these sums directly to spectral theory, they can deduce estimates which are much stronger than those which are obtained by using the Weil estimate term-by-term. Of course, this method requires explicit use of the spectra and estimates on the distribution of eigenvalues and the coefficients of the eigenforms.

As a specific example let me give the following result from **[15]**. For $T > 2$, $H = T^{2/3}$, and $\varepsilon > 0$,

$$\int_T^{T+H} |\zeta(1/2 + it)|^4 \, dt \ll HT^{\varepsilon}.$$

This improves the exponent from 7/8 to 2/3.

EXAMPLES 5. Sarnak **[27]** exploits the connection between the norms of conjugacy classes and lengths of geodesics to attack the class number problem for real quadratic fields. He recognizes that the norms of primitive hyperbolic conjugacy classes of $\Gamma(1)$ are exactly ε_d^2 where d is a discriminant number and that these appear with multiplicity $h(d)$. In other words, the lengths of closed geodesics on $\Gamma\backslash H$ are the numbers $2\log \varepsilon_d$. Then using the spectral theory to examine the distribution of closed geodesics, he deduces the following result.

$$\sum_{\varepsilon_d < x} h(d) = Li(x^2) + O(x^{3/2}(\log x)^2), \qquad x \to \infty.$$

One should compare this with Gauss' theorem **[3]**.

EXAMPLE 6. In **[11]** and **[12]**, I showed that the Dirichlet series (4) has the property that a positive proportion of its zeros lie on its critical line. The basic proofs follow along the lines of either Titchmarch **[32]** or Selberg **[28]**, with one major problem which needs to be treated in the context of the ideas described above. In either method, sums of the type

$$\sum_n \tau(n)\, \tau((bn + k)/a)$$

had to be estimated non-trivially and uniformly in the parameters a, b and k. This is like the problem (3) above with the additional difficulty of uniformity. I dealt with this question in **[10]** by studying the convolution of cusp forms on different groups. This required spectral estimates exhibiting the dependence on the group.

It should be mentioned that this theorem on the critical zeros holds for any cusp form on any congruence group which is also an eigenfunction of the appropirate Hecke operators. (In these papers, this result was only stated for $\Gamma(1)$.) It may even apply to those with multiplier system, as well.

In conclusion, I hope you are convinced of the depth and power of the interrelationships described in the first part of this paper. I also hope that some of this is a bit less mysterious to those unfamiliar with the spectral connection and perhaps have encouraged a few to begin a study of this rich and potentially lucrative subject.

REFERENCES

1. T. M. Apostol, *Modular Functions and Dirichlet Series in Number Theory*, Springer-Verlag, New York, 1976.

2. R. W. Bruggeman, *Kuznietsov's proof of the Ramanujan-Petersson conjecture for modular forms of weight zero*, Communications of the Math. Institute Rijksuniversiteit Utrecht **8** (1979).

3. C. F. Gauss, *Disquistiones Arithmeticae*, 304.

4. I. M. Gel'fand, M. I. Graev, and I. I. Piatetskii-Shapiro, *Representation Theory and Automorphic Functions*, Saunders Company, Philadelphia, 1969.

5. A. Good, *Beitraege zur Theorie der Dirichletreihen, die Spitzenformen zugeordnet sind*, J. Number Theory, **13** (1981), 18–65.

6. ——— *Cusp forms and eigenfunctions of the Laplacian*, Math. Ann. **255** (1981), 523–548.

7. ———, *The square mean of Dirichlet series associated to cusp forms*, preprint (1981).

8. ———, *On various means involving the Fourier coefficients of cusp forms*, preprint (1982).

9. ———, *New results on Ramanujan's function*, preprint (1982).

10. J. L. Hafner, *Explicit estimates in the arithmetic theory of cusp forms and Poincare series*, Math. Ann. (1983), to appear.

11. ———, *Zeros on the critical line for Dirichlet series attached to certain cusp forms*, Math. Ann. (1983), to appear.

12. ———, *On the zeros (a la Selberg) of Dirichlet series attached to certain cusp forms*, Proc. of the Summer Conference on Analytic Number Theory at the University of Texas at Austin, 1982, University of Texas Press, to appear.

13. D. A. Hejhal, *The Selberg Trace Formula and the Riemann zeta function*, Duke Math. Journal **43** (1976), 441–482.

14. H. Iwaniec, *On the mean values for Dirichlet polynomials and the Riemann zeta-function*, J. London Math. Soc. **22** (1980), 39–45.

15. ———, *Fourier coefficients of cusp forms and the Riemann zeta-function*, Sem. Th. Numbers, Bordeaux, (1979–1980), expose No. 18, 36 pp.

16. ———, *Mean values for Fourier coefficients of cusp forms and sums of Kloosterman sums*, Journée Arithmetiques de Exeter, 1980, Cambridge University Press, 1982.

17. ———, *On the Brun-Titchmarsh theorem*, J. Math. Soc. Japan, (1) **34** (1982), 95–122.

18. ———, *Promenade along modular forms and analytic number theory*, Proc. of the Summer Conference on Analytic Number Theory at the University of Texas at Austin, 1982, University of Texas Press, to appear.

19. T. Kubota, *Elementary Theory of Eisenstein Series*, John Wiley and Sons, New York, 1973.

20. N. V. Kuznetsov, *Petersson hypothesis for forms of weight zero and Linnik hypothesis (Russian)*, Preprint no, 2, Khab. K.H.I.I., Khabarovsk (1977).

21. ———, *Petersson hypothesis for parabolic forms of weight zero and Linnik hypothesis*. Sums of Kloosterman sums, Math. Sbornik 111 **153** (1980), 334–383.

22. J. Lehner, *Discontinuos Groups and Automorphic functions*, Amer. Math. Soc. Surveys, No. 8, 1964.

23. A. Ogg, *Modular Forms and Dirichlet Series*, W. A. Benjamin, Inc., New York, 1969.

24. B. Randol, *Small eigenvalues of the Laplace operator on compact Riemann surfaces*, Bull. Amer. Math. Soc. **80** (1974), 996–1000.

25. R. A. Rankin, *Modular Forms and Functions*, Cambridge University Press, 1977.

26. ———, *Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions*, I–III, Proc. Cambridge Phil. Soc. **35** (1939), 351–372; **36** (1940), 150–151.

27. P. Sarnak, *Class numbers of indefinite binary quadratic forms*, J. Number Theory **15** (1982), 229–247.

28. A. Selberg, *On the zeros of Riemann's zeta-function*, Skr. Norske. Vid.-Akad. Oslo I **10** (1942), 59pp.

29. ———, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. **20** (1956), 47–87.

30. ———, *On the estimation of Fourier coefficients of modular forms*, Proc. Symposia in Pure Math., VIII, A.M.S., (1965), 1–15.

31. J. G. Thompson, *Some finite groups which appear as* Gal L/K, *where* $K \subseteq Q(\mu_n)$, preprint (1982).

32. E. C. Titchmarsh, *The Theory of the Riemann Zeta Function*, Oxford University Press (Clarendon), London/New York, 1951.

33. A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U.S.A. **34** (1948), 204–207.

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA CA 91125
UNIVERSITY OF CALIFORNIA, SAN DIEGO, LA JOLLA, GA 92093