ELEMENTARY AND ANALYTIC THEORY OF NUMBERS
BANACH CENTER PUBLICATIONS, VOLUME 17
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

INTRODUCTION TO KLOOSTERMANIA

M. N. HUXLEY

Department of Pure Mathematics, University College
Cardiff, Great Britain

Contents

1. The Epstein zeta function
2. Hyperbolic space
3. The Kuznetsov Trace Formulae
4. Nodal line theorems
5. The group $\Gamma^0(13)$
6. Numerical bounds for the first eigenvalue
7. Exceptional eigenvalues in genus zero
8. The spectral theorem
9. Bessel functions
10. Fourier coefficients of cusp forms
11. Hecke operators
12. Kuznetsov's Basic Lemma
13. The First Trace Formula
14. The Second Trace Formula
15. Linnik's Sum
16. Kloosterman sums with opposite signs

General notation

ch u (pronounced 'kosz') is cosh u (cosinus hyperbolicus).
cis x (pronounced 'kis') is $\cos x + i \sin x$.
$d(n)$ is the ordinary divisor function.
$d_3(n)$ is $\sum_{m|n} d(m)$.
$e(x)$ is $e^{2\pi i x}$.
exp x is written instead of e^x when x is a complicated expression.
i is the square root of -1. Kubota uses $\sqrt{(-1)}$.
Kh y is the Hänkel function of pure imaginary argument. Kuznetsov uses $K(y)$ and normalises differently.

$\underline{|s}$ is the factorial of s. Kuznetsov uses $\Gamma(s+1)$ or $s!$
$\log x$ is the logarithm to base e. Kuznetsov uses $\ln x$.
$\operatorname{sh} u$ (pronounced 'sańś') is $\sinh u$ (sinus hyperbolicus).
$\tau_s(n)$ is a divisor function $\sum_{de=n} \left(\frac{d}{e}\right)^s$.
$\operatorname{th} u$ (pronounced 'tansz') is $\tanh u$ (tangens hyperbolica), $\sinh u/\cosh u$.

1. The Epstein zeta function

Additive problems in number theory are usually attacked using

$$e(x) = \exp 2\pi i x,$$

and multiplicative ones using n^{it}, or $\chi(n)$ if there is a congruence condition. Kuznetsov, Iwaniec and Deshouillers have added to this list. I will try to explain how their new functions arise.

The basic lemma in number theory, Euclid's algorithm, states that given two integers a and c with $(a, c) = 1$, there are integers b, d with

$$ad - bc = 1.$$

This condition says that an integer matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

lies in $\mathrm{SL}_2(Z)$, the group of integer matrices of determinant one. The new functions associated with it are called Eisenstein series and Maass wave forms.

The function $e(x)$ has $e(x+n) = e(x)$. An analogue for matrices would have

$$f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right) = f(z).$$

We have not yet defined z. What do 2×2 matrices act on? Here Bruggeman would say, on themselves by matrix multiplication. But there is a simpler action. Let z be a complex number,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}.$$

This product is associative, in the sense that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} r & s \\ t & u \end{pmatrix} z = \begin{pmatrix} ar+bt & as+bu \\ cr+dt & cs+du \end{pmatrix} z.$$

When we allow ∞ as a value for z, the action makes sense for any real numbers a, b, c, d with $ad-bc>0$, with

$$\operatorname{Im} z>0,\ z\neq\infty \Rightarrow \operatorname{Im}\begin{pmatrix} a & b \\ c & d\end{pmatrix} z>0.$$

Hence we may restrict z to the upper half plane H defined by the condition $\operatorname{Im} z>0$.

Do such functions f exist? Yes. Consider the Epstein zeta function of the lattice generated by 1 and z, where $\operatorname{Im} z>0$,

$$B(z, s)=\sum_{\substack{c=-\infty \\ (c,d)\neq(0,0)}}^{\infty}\sum_{d=-\infty}^{\infty}\frac{y^s}{|cz+d|^{2s}}.$$

The function $B(z, s)$ depends only on the lattice, and it satisfies

$$B(z, s)=B(z+1, s)=B(-\bar{z}, s)=B(-1/z, s),$$

and by induction

$$B\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix} z, s\right)=B(z, s)$$

for any matrix in $\mathrm{SL}_2(Z)$. So without loss of generality we may suppose

$$|z|\geqslant 1,\ y>0,\ -1/2\leqslant x\leqslant 1/2.$$

These conditions define the fundamental domain D (Fig. 1).

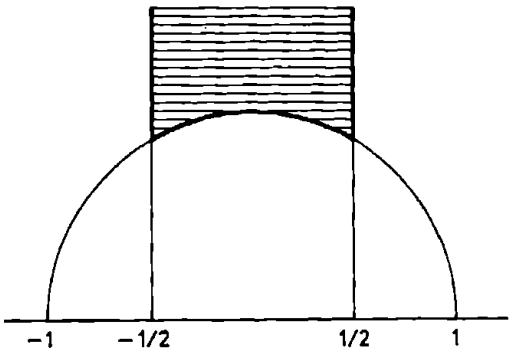

Fig. 1

When z is restricted as above, the lattice point argument of Vinogradov or van der Corput gives

$$N(r):=\sum_{0<|cz+d|\leqslant r} 1=\frac{\pi r^2}{y}+O(r^{2/3}y^{2/3}\log ry),$$

so we obtain

$$B(z, s) = sy^s \int_1^\infty \frac{N(\sqrt{t})}{t^{s+1}}\, dt = \frac{\pi s}{s-1}\, y^{s-1} + R(s),$$

where $R(s)$ is regular for $\operatorname{Re} s > 1/3$. We deduce that $B(z, s)$, which converges absolutely only for $\operatorname{Re} s > 1$, has a meromorphic continuation to $\operatorname{Re} s > 1/3$ with a pole at $s = 1$, residue π.

We can check that $B(z, s)$ satisfies the differential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) B(x+iy, s) = \frac{s(s-1)}{y^2}\, B(x+iy, s)$$

term by term. Also $B(z, s)$ must have a Fourier series in x, whose coefficients are functions of y, and these functions of y tend to infinity no faster than a fixed power of y (depending on s).

$$B(z, s) = \sum_{-\infty}^{\infty} a_n(y)\, e(nx),$$

with

$$y^2 a_n''(y) = \{4\pi^2 n^2 y^2 - s(1-s)\}\, a_n(y).$$

For $n = 0$ we know one solution must be y^s; the other is y^{1-s}. For $n \neq 0$ we may put

$$a_n(y) = a(2\pi y |n|),$$

and if $a(u) = \sqrt{u}\, f(u)$

$$(u^2 - s(1-s)) f(u) = u(uf')' - \tfrac{1}{4} f,$$

which is Bessel's equation, and has two linearly independent solutions, of which one increases exponentially in u. So we want the other one. Naming names,

$$B(z, s) \sim ay^s + by^{1-s} + 2 \sum_{n=1}^{\infty} c_n \cos 2\pi nx A_{s-1/2}(2\pi ny),$$

where $A_k(u)$ is expressed in terms of more familiar Bessel functions by

$$A_k(u) = \sqrt{\frac{\pi u}{2}}\, \mathrm{Kh}_k(u) = \sqrt{\frac{\pi u}{2}}\, \frac{I_{-k}(u) - I_k(u)}{\sin \pi k},$$

and

$$A_k(u) \sim e^{-u} \quad \text{as} \quad u \to +\infty.$$

Let us calculate a, b and c_n.
For $\operatorname{Re} s > 1$ the $n = 0$ term is

$$2\sum_1^\infty \frac{1}{d^{2s}}\, y^s + 2\int_0^1 \sum_{c=1}^\infty \sum_{d=-\infty}^\infty \frac{y^s}{\{(cx+d)^2 + c^2 y^2\}^s}\, dx.$$

So

$$a = 2\zeta(2s),$$

$$\tfrac{1}{2} b = y^{2s-1} \sum_{c=1}^\infty \sum_{d \bmod c} \int_{-\infty}^\infty \frac{dx}{\{(cx+d)^2 + c^2 y^2\}^s}$$

(grouping d into residue classes)

$$= y^{2s-1} \sum \frac{c}{c^{2s}} \int_{-\infty}^\infty \frac{dx}{(x^2+y^2)^s}$$

$$= \zeta(2s-1)\, 2\int_0^\infty \frac{dt}{(t^2+1)^s} = \zeta(2s-1) \int_0^\infty \frac{du}{\sqrt{u}\,(u+1)^s}$$

$$= \frac{\lfloor s-3/2 \;\; \lfloor -1/2}{\lfloor s-1}\, \zeta(2s-1),$$

where $\lfloor s$ is the factorial function $\Gamma(s+1)$, and the last step is Euler's trick evaluation of the Beta function integral. You will be pleased to hear that c_n is worse.

$$c_n A(2\pi n y) = 2\int_0^1 e(nx) \sum_{c=1}^\infty \sum_{d=-\infty}^\infty \frac{y^s}{\{(cx+d)^2 + c^2 y^2\}^{2s}}\, dx.$$

As before, we collect terms in d into residue classes mod c:

$$= 2\sum_{c=1}^\infty \frac{1}{c^{2s}} \sum_{d \bmod c} \int_{-\infty}^\infty \frac{y^s e(nx)}{\{(x+d/c)^2 + y^2\}^s}\, dx$$

$$= 2\sum_{c=1}^\infty \frac{1}{c^{2s}} \sum_{d \bmod c} e\left(\frac{-nd}{c}\right) \int_{-\infty}^\infty \frac{y^s e(nx)}{(x^2+y^2)^s}\, dx.$$

The sum over d is 0 unless $c \mid n$.

We know the integral over x is a multiple of $A(2\pi ny)$. Which?

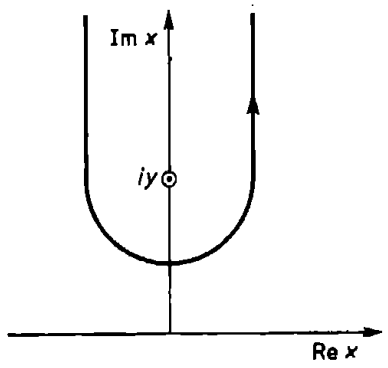

Fig. 2

Make it a contour integral, and deform the contour as shown in Fig 2. For large y only that part near iy matters, and it is asymptotically a Γ-integral, so that we have

$$\sim 2\sum_{c|n} c^{1-2s}\frac{(\pi n)^s}{\lfloor s-1}\frac{e^{-2\pi ny}}{n} \sim \frac{2\pi^s}{\lfloor s-1}\frac{\tau_{s-1/2}(n)}{\sqrt{n}}A(2\pi ny),$$

where for simplicity

$$\tau_u(n) = \sum_{de=n}\left(\frac{d}{e}\right)^u = \tau_{-u}(n).$$

So we have identified c_n, and

$$\frac{\lfloor s-1}{\pi^s}B(z,s) = \frac{2\lfloor s-1}{\pi^s}\zeta(2s)y^s + \frac{2\lfloor s-\frac{3}{2}}{\pi^{s-1/2}}\zeta(2s-1)y^{1-s} +$$

$$+4\sum_1^\infty \frac{\tau_{s-1/2}(n)}{\sqrt{n}}\cos(2\pi nx)A_{s-1/2}(2\pi ny).$$

This hard work had better be some use.

The Bessel function is defined for all s, and it is an even function of $s-1/2$, so the sum over n gives a regular function of s. For $1/2 < \mathrm{Re}\ s \leqslant 1$ all the terms are defined except $\zeta(2s-1)$. So $\zeta(s)$ must have a continuation to $0 < \mathrm{Re}\ s \leqslant 1$ with a pole at $s=1$ residue 1. Well there are easier ways to prove that. I claim next that

$$\frac{\lfloor s-1}{\pi^s}B(z,s) = \frac{\lfloor -s}{\pi^{1-s}}B(z,1-s)$$

since both sides are defined for $1/3 < \mathrm{Re}\ s < 2/3$. Their difference is a function of y only, satisfying the functional equations for $B(z,s)$, and zero at

$s = 1/2$. So the difference is 0. Comparing the coefficients of y^s

$$\frac{\left|\frac{1}{2}s-1\right.\zeta(s)}{\pi^{s/2}} = \frac{\left|-\frac{1}{2}-\frac{1}{2}s\right.\zeta(1-s)}{\pi^{1/2-s/2}}$$

and we have the functional equations for the Riemann zeta functions and the Epstein zeta functions.

Suppose now that $\zeta(1+i\gamma) = 0$ for some γ. Let $s_0 = \frac{1}{2}(1+i\gamma)$. The terms in y^{s_0} and y^{1-s_0} in the Fourier series for $B(z, s_0)$ vanish. In fact $B(z, s_0)$ will tend to zero exponentially as $y \to \infty$, becoming what is called a cusp form.

The Eisenstein series is defined for $\operatorname{Re} s > 1$ by taking out common factors between b and c in the definition of $B(z, s)$:

$$E(z, s) = \frac{B(z, s)}{2\zeta(2s)} = \tfrac{1}{2} \sum_{\substack{c=-\infty \\ (c,d)=1}}^{\infty} \sum_{d=-\infty}^{\infty} \frac{y^s}{|cz+d|^{2s}}.$$

Writting $B(z)$ for $B(z, s_0)$ we have

$$\int\int_D E(z, s)\overline{B(z)}\,\frac{dx\,dy}{y^2} = \tfrac{1}{2} \sum_{\substack{c=-\infty \\ (c,d)=1}}^{\infty} \sum_{d=-\infty}^{\infty} \int\int_D \frac{y^s}{|cz+d|^{2s}}\,\overline{B(z)}\,\frac{dx\,dy}{y^2}.$$

Now the sum over c and d is in (1, 1) correspondence with the cosets of the subgroup generated by $z \to z+1$ in the action of $\mathrm{SL}_2(Z)$ on H. This action is called $\mathrm{PSL}_2(Z)$ and differs from $\mathrm{SL}_2(Z)$ only in that $-I$ acts as the identity. Hence the integral is

$$\sum_{\text{cosets}} \int\int_D \left\{\operatorname{Im}\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right\}^s \overline{B(z)}\,\frac{dx\,dy}{y^2}$$

$$= \sum_{\text{cosets}} \int\int_{\binom{d}{-c}\ \binom{-b}{a}D} y^s\overline{B(z)}\,\frac{dx\,dy}{y^2} = \int_{y=0}^{\infty} \int_{x=-1/2}^{1/2} y^s\overline{B(z)}\,\frac{dx\,dy}{y^2}.$$

This integral is now zero, since every term in the Fourier series for $B(z)$ contains an $e(nx)$ with $n \neq 0$.

To obtain a contradiction, we note that since $\zeta(2s_0) = 0$,

$$\zeta(2s) \sim a(s-s_0)^r$$

for some $r \geqslant 1$ and some nonzero a. Take a contour enclosing s_0 and no other pole of $E(z, s)$. Since the triple integral involved converges absolutely,

we have

$$\frac{1}{2\pi i}\oint_C (s-s_0)^{r-1}\int\int_D E(z,s)\overline{B(z,s_0)}\,\frac{dx\,dy}{y^2}\,ds$$

$$=\int\int_D \frac{1}{2\pi i}\oint_C (s-s_0)^{r-1}E(z,s)\,ds\,\overline{B(z,s_0)}\,\frac{dx\,dy}{y^2}$$

$$=\int\int_D \frac{B(z,s_0)\overline{B(z,s_0)}}{2a}\,\frac{dx\,dy}{y^2}=\frac{1}{2a}\int\int_D |B(z,s_0)|^2\,\frac{dx\,dy}{y^2}.$$

Since $B(z,s_0)$ is not identically zero (consider large y), this integral is nonzero. However

$$\int\int_D E(z,s)\overline{B(z,s_0)}\,\frac{dx\,dy}{y^2}$$

is the analytic continuation of a function of s which is identically zero for $\operatorname{Re} s>1$, and so it remains zero for all s. This contradiction now gives $\zeta(1+i\gamma)\neq 0$.

Background reading
JEFFREYS and JEFFREYS
KUBOTA
VINOGRADOV
ZAGIER

2. Hyperbolic space

You will now ask: what sort of function is this Eisenstein or Epstein series? Consider it for fixed s, so it now becomes a function of z, with

$$E(\gamma z,s)=E(z,s)\quad\text{for }\gamma=\begin{pmatrix}a&b\\c&d\end{pmatrix}\quad\text{in }\Gamma(1)=\mathrm{PSL}(2,Z),$$

$$y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)E(z,s)=-s(1-s)E(z,s).$$

The first condition says that we only need take z in $|z|\geqslant 1$, $-1/2\leqslant x\leqslant 1/2$, $y>0$, the fundamental region D on the upper half plane. The second condition says that $E(z,s)$ is an eigenfunction of the Laplacian. Let me explain what this means.

The simplest model of the hyperbolic plane is the upper half plane H, with circles orthogonal to the x-axis representing straight lines. The two

intersections of such circles are on opposite sides of the x-axis, so two straight lines meet in at most one point. The angle between two straight lines is what it looks like, the angle between their tangents. Given two points A, B, the circle through A and B with centre on the x-axis is unique. It cuts the x-axis in two points U, V (Fig. 3).

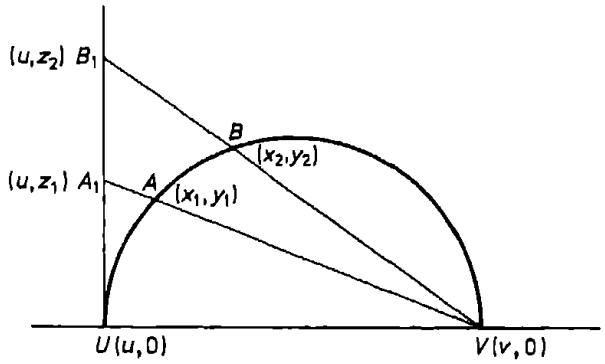

Fig. 3

The distance from A to B is defined as the logarithm of the cross-ratio $d = \log (AB, UV)$. Draw lines from V to A, B, U, V, cutting the tangent at U in A_1, B_1, U, V_1. V_1 is at infinity, so off the picture. Then

$$(AB, UV) = (A_1 B_1, UV_1) = z_1/z_2.$$

There are various equal ratios, since $UA \perp AV$, $UB \perp BV$

$$\frac{z_1}{u-v} = \frac{y_1}{x_1-v} = \frac{-(x_1-u)}{y_1}, \quad \frac{z_2}{u-v} = \frac{y_2}{x_2-v} = \frac{-(x_2-u)}{y_2}.$$

Hence

$$e^d = \frac{z_1}{z_2} = \frac{-(x_1-u)(x_2-v)}{y_1 y_2} = \frac{-y_1 y_2}{(x_2-u)(x_1-v)}.$$

We eliminate u and v by symmetry:

$$2 \operatorname{ch} d = \frac{z_1}{z_2} + \frac{z_2}{z_1} = \frac{-\{2x_1 x_2 - (u+v)(x_1+x_2) + 2uv\}}{y_1 y_2}$$

$$= \frac{(x_1-x_2)^2 - (x_1-u)(x_1-v) - (x_2-u)(x_2-v)}{y_1 y_2} = \frac{(x_1-x_2)^2 + y_1^2 + y_2^2}{y_1 y_2},$$

which is usually written as

$$4 \operatorname{sh}^2 \frac{d}{2} = 2 \operatorname{ch} d - 2 = \frac{|z_1 - z_2|^2}{y_1 y_2}.$$

Infinitesimally the metric is

$$ds^2 = \frac{dx^2+dy^2}{y^2},$$

area is given by

$$d\mu = \frac{dx\,dy}{y^2}.$$

The hyperbolic Laplacian must transform contragredient to the area, so it is

$$y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right).$$

If a, b, c, d are real, $z \to \frac{az+b}{cz+d}$ preserves cross ratios, so the metric is invariant under the action of SL(2, R). In fact, $z \to \frac{az+b}{cz+d}$ gives all the rigid motions which do not involve reflection.

Straight lines and circles in the Euclidean metric correspond to lines in the hyperbolic metric as follows.

1. Euclidean circles or straight lines perpendicular to the real axis represent hyperbolic straight lines.

2. Euclidean circles not cutting the real axis represent hyperbolic circles (locus of a point a fixed distance from a fixed point.)

3. Euclidean circles touching the x-axis or straight lines $y =$ constant are called horocycles.

4. Euclidean circles or straight lines cutting the real axis at an acute angle are hypercycles-locus of a point a fixed distance from a fixed line.

From now on 'straight line' will mean an hyperbolic straight line. Rigid motions can be classified as

1. Translation along a straight line. Points not on the straight line move along hypercycles. (Russian: giperbolicheskaya)

2. Rotations. One point fixed, all others move in circles. (Russian: ellipticheskaya)

3. Transvections (or rotations around points at infinity). One direction (or one point at infinity) fixed, all finite points move along horocycles. (Russian: parabolicheskaya)

Some useful theorems and formulae:

GAUSS'S DEFECT THEOREM. *The area of an n-sided polygon and its internal angles add up to* $(n-2)\pi$.

Proof (Fig. 4). The area is $\int\int \frac{dx\,dy}{y^2}$. Integrating over y, this is

$$\int_{\text{lower sides}} \frac{dx}{y} - \int_{\text{upper sides}} \frac{dx}{y}.$$

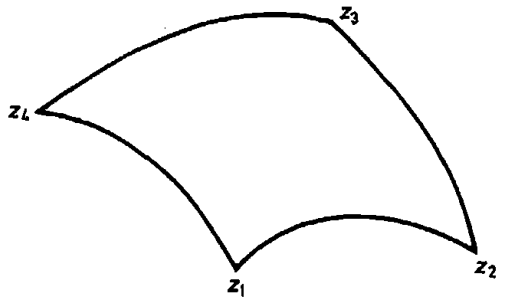

Fig. 4

The sides are circular arcs, and

$$\int_{\beta}^{\alpha} \frac{d(r\cos\theta)}{r\sin\theta} = \beta - \alpha.$$

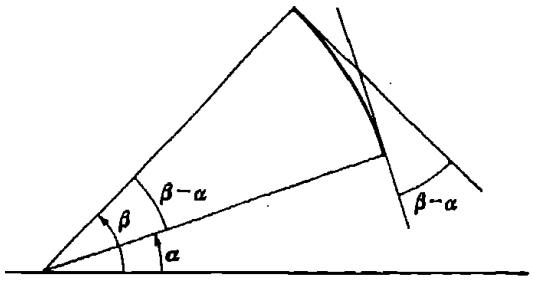

Fig. 5

The contribution of one side to the angle-defect $(n-2)\pi$-sum of angles is $\beta-\alpha$, positive for lower sides, negative for upper sides.

COROLLARY. *The area of quadrilateral* $z_1 z_2 z_3 z_4$ *is*

$$\text{Im log}\,(z_1, z_3; \bar{z}_2, \bar{z}_4).$$

A question: If the lengths of sides of a quadrilateral are fixed, is the area greatest when the vertices are on a circle? This is true in the Euclidean plane, Zenodorus's theorem. It would give a nice proof that of all closed curves of

given length, the circle has the greatest area, a fact I need in a later lecture. If Zenodorus's theorem is true in hyperbolic geometry, it ought to be well-knowable.

Some formulae: lines are best described by their endpoints on the real axis.

Distance of z from line (a, b):

$$\operatorname{ch} d = \frac{|z-b|\,|z-a|}{y\,|a-b|}.$$

Distance of line (m, n) from line (r, s) (if non intersecting):

$$\coth^2 \frac{d}{2} = (m, n; r, s).$$

Angle between lines (m, n) and (r, s) (if intersecting):

$$\tan^2 \frac{\theta}{2} = -(m, n; r, s).$$

Trigonometry:

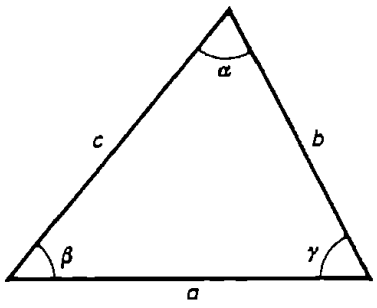

Fig. 6

Cosine Rule: $\operatorname{ch} c = \operatorname{ch} a \operatorname{ch} b - \operatorname{sh} a \operatorname{sh} b \cos \gamma$,

Sine Rule: $\dfrac{\sin \alpha}{\operatorname{sh} a} = \dfrac{\sin \beta}{\operatorname{sh} b} = \dfrac{\sin \gamma}{\operatorname{sh} c}$,

Cosh Rule: $\cos \alpha \cos \beta + \cos \gamma = \sin \alpha \sin \beta \operatorname{ch} c$, all fairly easy to check.

The circle radius ϱ has area $2(\operatorname{ch} \varrho - 1)$, circumference $2\pi \operatorname{sh} \varrho$. You can see why it is called hyperbolic geometry.

Note that most of the area inside a circle is close to the boundary. I find this very hard to visualise. It makes lattice point problems in hyperbolic space very difficult and counter-intuitive.

Now suppose you take the hyperbolic plane and hit it with a hammer. Think of it as an infinite thin sheet of metal. If you hit it simultaneously at all points P which are images of a fixed point P_0 under $\Gamma(1)$, it will vibrate

with the symmetry of $\Gamma(1)$. If you are lucky you will hear a pure musical note. That means you have set it into a normal mode of vibration.

The Eisenstein series $E(z, s)$ gives a normal mode of vibration

$$E(z, s) \cos nt, \quad n = c\sqrt{\lambda} = c\sqrt{s(1-s)}.$$

This is technically in the continuous spectrum, which means it has infinite energy per unit area, so it is difficult to obtain in practice. You can get an average over s of the Eisenstein series. The normal modes which carry finite energy per unit area are called square integrable eigenfunctions. All but finitely many of these die out at infinity – strictly towards $i\infty$. One that does not is the trivial eigenfunction where the whole plane is moved bodily downwards and then stops still in its new position. This one has eigenvalue 0, so you cannot hear it. The non-trivial normal modes of finite energy are called the Maass wave forms.

3. The Kuznetsov Trace Formulae

So far we have produced the Epstein or Eisenstein series, which is a very elaborate generating function for the divisor function. The aim of this theory is to generate Kloosterman sums.

In the Euclid's algorithm equation $ad-bc = 1$, it would often help to know how d varies with a for fixed c, or how d varies with c for fixed a. The method of Kloosterman and Hooley leads to the sum

$$S(m, n; c) = \sum_{\substack{a \bmod c \\ ad \equiv 1 (\bmod c)}} \sum_{d \bmod c} e\left(\frac{am+dn}{c}\right).$$

The individual terms of the Kloosterman sum are functions of 2×2 matrices, and the Kloosterman sum arises naturally in the Fourier coefficients of functions "on D" (by which we mean invariant under $\Gamma(1) = \mathrm{PSL}_2(Z)$). Put $h(z) = e(mx)g(y)$ and consider the sum

$$\sum h\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} z\right)$$

extended over a representative set of 2×2 matrices. This is Poincaré's construction; the case $m = 0$, $g(y) = y^s$ gives the Eisenstein series or Epstein zeta function. Since

$$\frac{az+b}{cz+d} = \frac{a}{c} - \frac{1}{c(cz+d)},$$

it is easy to see where the factor $e(an/c)$ comes from. In the formula for the nth Fourier coefficient there is a factor $e(-nx)$. There is a variable change

that replaces z by $z-d/c$, which gives

$$\sum e\left(\frac{ma}{c}\right)e\left(\frac{nd}{c}\right)\int_{-\infty}^{\infty} e\left(-m\,\mathrm{Re}\,\frac{1}{c^2 z}\right)g\left(\mathrm{Im}\,\frac{1}{c^2 z}\right)e(-nx)\,dx,$$

after we have put all the terms together with d in the same residue class (mod c). You can now see the Kloosterman sum appearing.

The idea of Kuznetsov and Bruggeman is that

$$U(z)=\sum h\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}z\right)$$

is a mode of vibration of the hyperbolic plane with the symmetry of $\Gamma(1)$, and it can be analysed into Eisenstein series and Maass wave forms. More precisely, they take two Poincaré series of this type, and calculate the energy per unit area of the mode of vibration $U_1(z)+tU_2(z)$, concentrating on the term which is linear in t. The calculation can be done in two ways. The 'unfolding method' sketched in the first section gives the energy in terms of the m_1th Fourier coefficient of U_2, an explicit but messy computation. On the other hand, the energy is the sum of the energies in the normal modes. Kuznetsov was able to choose a function $g(y)$ for which these integrals came out not in closed form, but in an usable one. This equality is the Basic Lemma of Kuznetsov's method. Next he picks an arbitrary analytic function, and integrates to obtain the First Trace Formula, which I so name by analogy with Selberg's Trace Formula. It contains on the left the Fourier coefficients of the Maass wave forms, which unfortunately are completely unknown, and on the right a sum over Kloosterman sums.

At this stage there are two cases. If m and n are both positive, one can choose a function which produces on the right-hand side the Kloosterman sum zeta function

$$Z_{m,n}(s)=\sum_1^{\infty}\frac{S(m,n;c)}{c^{2s}},$$

and on the left-hand side a convergent series of factorial functions. From this we read off the poles and residues of $Z_{m,n}(s)$. There are rather a lot of these, as it is an analogue of

$$\frac{1}{\zeta(s)}\,\frac{1}{\zeta(s+1)}\,\frac{1}{\zeta(s+2)}\cdots$$

(Kloosterman sums are associated with the Möbius function). The poles are at

(i) $s(1-s)=\lambda_n$ the eigenvalue of a normal mode,

(ii) $\frac{1}{2}\varrho$ where ϱ is a zero of $\zeta(s)$,

(iii) $\varkappa - 1$ if $\varkappa$ is some other pole (now iterate).

The residues are explicit, except that they contain the Fourier coefficients $b_j(n)$ of the Maass wave forms

$$f_j(z) \sim \sum_{-\infty}^{\infty} b_j(n)\, e(nx)\, A(2\pi n y).$$

One can now do a Mellin transform integral to pick out the coefficients $S(m, n; c)$ of $Z_{m,n}(s)$. There are lots of poles, so we need a weight function $g(c)$ whose Mellin transform $G(s)$ falls off like $1/t^2$. This gives Kuznetsov's Second Trace Formula:

$$\frac{1}{2\pi\sqrt{mn}} \sum_c g\left(\frac{c}{2\pi\sqrt{mn}}\right) S(m, n; c)$$

$$= \sum_{j=1}^{\infty} \frac{4\sqrt{mn}\ \overline{b_j(m)}\, b_j(n)}{\operatorname{ch} \pi\varkappa_j}\, \hat{g}(\varkappa_j) + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\tau_{it}(m)\, \tau_{it}(n)\, g(t)}{\zeta(1+2it)\, \zeta(1-2it)}\, dt -$$

$$- \sum_k \frac{(-1)^k}{2\pi} \sqrt{\left(\frac{m}{n}\right)^{2k-1}} \{c_n(G_{2k,n}) - \delta_{mn}\}(2k-1) \int J_{2k-1}(2/x)\, g(x)\, dx.$$

Here the extra notation is $\lambda_j = \frac{1}{4} + \varkappa_j^2$, $c_n(G_{2k,m})$ is the coefficient of a classical modular form which arises from the trivial zeros of $\zeta(s)$ in order to satisfy Bruggeman's Philosophy, and $\hat{g}$ ought to be called Kuznetsov's transform

$$\hat{g}(t) = -\frac{\pi}{\operatorname{sh} \pi v} \int \operatorname{Im} J_{2it}(2/x)\, g(x)\, dx.$$

Thus if $g(x)$ is the piecewise linear function whose graph with $\Delta > 1$ is

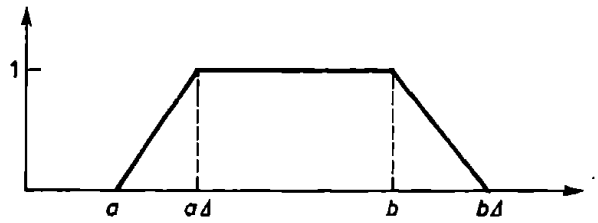

Fig. 7

then

$$\sum_q g\left(\frac{q}{2\pi\sqrt{mn}}\right) S(m, n; q) \ll \frac{N}{\sqrt{\Delta - 1}} \left\{1 + \sqrt{\Delta - 1}\left(1 - \frac{M}{N}\right) d(n) \sqrt{n} \log^2 n\right\}^{1/2} \times$$

$$\times \left\{1 + \sqrt{\Delta - 1}\left(1 - \frac{M}{N}\right) d(m) \sqrt{m} \log^2 m\right\}^{1/2},$$

with $M = 2\pi a\sqrt{mn}$, $N = 2\pi b\sqrt{mn}$.

Many more elaborate means are in the paper of Deshouillers and Iwaniec [9].

The case $m > 0$, $n < 0$ differs. The sum over residues of $Z_{m,n}(s)$ does not converge, but the Second Trace Formula comes out by a different argument, with a very delicate inversion in *it*. The Kuznetsov transform is replaced by a Lebedev–Kontorovich transform, and there is no term in the classical modular forms.

To deal with congruence conditions in ordinary multiplicative number theory we use the characters $\chi(n)$ as well as n^{it}. In this theory, to sum c over multiples of some fixed Q, we go to the congruence subgroup $\Gamma^0(Q)$ of the full modular group $\Gamma(1)$. Since we require less symmetry, there are many new normal modes. When Q is squarefree there are no essentially new Eisenstein series, but for Q with a squared factor we get the Epstein zeta function of a shifted lattice.

The difficulty arises from the cusp forms. For $\lambda \geqslant 1/4$ the eigenvalue of a Maass wave form is written as

$$\lambda_j = \tfrac{1}{4} + \varkappa_j^2, \qquad \varkappa_j \geqslant 0.$$

For $\lambda < 1/4$ we must write

$$\lambda_j = \beta_j(1-\beta_j), \qquad 1/2 < \beta_j \leqslant 1.$$

For $\Gamma(1)$ we have $\lambda_1 = 91.1413\ldots$, but eigenvalues $< 1/4$ are not impossible on theoretical grounds. The number β_j is analogous to the possible exceptional zero of a real quadratic L-function, so that

$$\sum_{q \equiv 0 (\mathrm{mod}\, Q)} g\left(\frac{q}{2\pi\sqrt{mn}}\right) S(m, n; q) \sim cN^{\beta_j + 1/2},$$

where the constant c depends on m and n as well as on the weight function g.

In the next four sections of this paper we use combinatorial geometry to discuss possible small eigenvalues, in the spirit of Buser.

Background reading

BRUGGEMAN
DESHOUILLERS–IWANIEC
HOOLEY
KUZNETSOV
MAASS
MAGNUS
RANKIN

4. Nodal line theorems

In this section results on partial differential equations are sketched. An *eigenfunction* means a function on the upper half plane H which satisfies

$$\text{div grad } f = -\lambda f$$

in the hyperbolic metric. A *domain* will mean the inside of a smooth simple closed curve. We say that a function g satisfies *boundary conditions on a domain* D if one of the following conditions holds:

(i) $g = 0$ on the boundary ∂D (Dirichlet condition).

(ii) The gradient vector grad g points along the boundary, so that the outward normal derivative satisfies

$$\frac{\partial g}{\partial n} = 0 \quad \text{on } \partial D \quad \text{(Neumann condition)}.$$

(iii) The boundary ∂D can be divided into finitely many arcs which are congruent in pairs, and g takes the same value at corresponding points (periodic condition).

(iv) The boundary ∂D can be divided into finitely many arcs, on each of which one of the conditions (i), (ii) or (iii) holds.

The boundary should be sufficiently smooth for the divergence theorem to hold for a sufficiently differentiable vector $\boldsymbol{u}$

$$\oint_{\partial D} u_n \, ds = \int\int_D \text{div}\, \boldsymbol{u} \, d\mu,$$

where u_n denotes the component of the vector $\boldsymbol{u}$ in the direction of the outward normal. When f is a nontrivial eigenfunction the contour lines of f, the sets $f^{-1}(t)$ for some constant t, form smooth curves.

If now g is a function satisfying boundary conditions on D,

$$\oint_{\partial D} g \frac{\partial g}{\partial n} \, ds = 0,$$

and expanding div (g grad g) we have from the divergence theorem

$$\int\int_D |\text{grad } g|^2 \, d\mu = -\int\int_D g \text{ div grad } g \, d\mu.$$

In particular, for an eigenfunction f satisfying boundary conditions on D we have

$$\int\int_D |\text{grad } f|^2 \, d\mu = \lambda \int\int_D f^2 \, d\mu.$$

We introduce more terminology. We say f is an *eigenfunction on* D if

(i) div grad $f = -\lambda f$ on D,
(ii) f satisfies boundary conditions on ∂D,
(iii) $\int\int_D f^2 \, d\mu$ converges.

With this convention the Eisenstein series is not an eigenfunction on D when D is the fundamental region for the full modular group, but it may be an eigenfunction on E for a subset E of D which does not include the point at infinity ("cusp") $i\infty$.

A *nodal line* of a function g is a curve on which $g = 0$, and a *nodal domain* is the interior of a smooth simple closed nodal line. A nodal domain is *minimal* if no proper subset of it is a nodal domain. We show first that an eigenfunction f has constant sign on its minimal nodal domains, and the nodal domains determine the eigenfunction, in the sense that two linearly independent eigenfunctions cannot have a common minimal nodal domain.

For fixed domain and boundary conditions, we number the eigenvalues according to multiplicity $\lambda_1, \lambda_2, \ldots$ and corresponding eigenfunctions, orthogonal in the sense that

$$\int\int_D f_i f_j \, d\mu = \delta_{ij},$$

are numbered $f_1, f_2, \ldots$ If a constant function satisfies the boundary conditions, it will be f_1, with corresponding $\lambda_1 = 0$. In other chapters the zero eigenvalue is λ_0, but to name it so here would upset the counting arguments. We make the further convention that only eigenfunctions with eigenvalues below the 'continuous spectrum' are counted. For our application this means that if the Eisenstein series satisfy the boundary conditions on D, then we count only exceptional eigenvalues, those with $\lambda < 1/4$. We can now state the nodal line theorems.

HOPF'S RULE. *A nonconstant function g with* div grad $g \leqslant 0$ *cannot have a local minimum.*

Proof. Take a circle C with centre P, bounding a set E.

$$\oint_C \frac{\partial g}{\partial n} \, ds = \int\int_E \text{div grad } g \, d\mu \leqslant 0.$$

If P was a local minimum, there would be a small circle C with $\partial g/\partial n \geqslant 0$.

COROLLARY 1. *An eigenfunction changes sign across a nodal line.*

COROLLARY 2. *A nodal line of an eigenfunction which enters a domain D divides D into two non empty sets, and so must cut ∂D again.*

COROLLARY 3. *An eigenfunction with $f \geqslant 0$ on D has $f > 0$ on D.*

A TEST FOR THE FIRST EIGENFUNCTION. *If f is an eigenfunction on D with $f > 0$ on D, then f minimises the ratio*

$$\int\int_D |\text{grad } g|^2 \, d\mu / \int\int_D g^2 \, d\mu$$

among all functions satisfying the boundary conditions.

Proof. By the boundary conditions

$$\oint_{\partial D} \frac{g^2}{f} \frac{\partial f}{\partial n} \, ds = 0.$$

Applying the divergence theorem, we see that

$$0 = \int\int_D \frac{g^2}{f} \text{ div grad } f \, d\mu + \int\int_D \left(\frac{2g}{f} \text{ grad } g \cdot \text{grad } f - \frac{g^2}{f^2} |\text{grad } f|^2 \right) d\mu$$

$$= -\lambda \int\int_D g^2 \, d\mu + \int\int_D \left| \text{grad } g - \frac{g}{f} \text{ grad } f \right|^2 d\mu + \int\int_D |\text{grad } g|^2 \, d\mu,$$

which gives the bound, with equality only if g is proportional to f.

COROLLARY. *If f is an eigenfunction on D with $f > 0$ on D, f must be the first eigenfunction.*

COURANT'S NODAL LINE THEOREM. *The n-th eigenfunction on D has at most n minimal nodal domains. If there are periodic boundary conditions, two or more nodal domains which become connected if we identify the corresponding arcs of the boundary are counted as one nodal domain.*

The counting convention under periodic boundary conditions is best explained with a diagram:

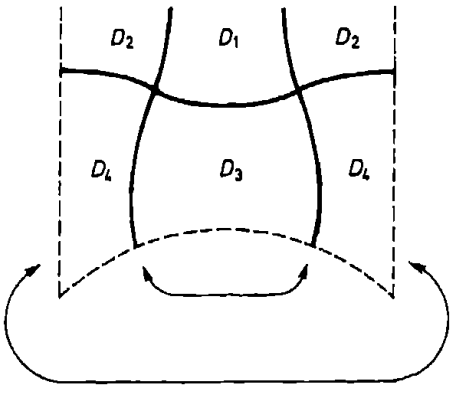

Fig. 8

Proof. Let f be an eigenfunction on D with nodal domains $D_1, \ldots, D_k$. Let χ_i be the characteristic function of D_i and choose real constants $a_1, \ldots, a_{k-1}$ so that

$$g = (a_1 \chi_1 + \ldots + a_{k-1} \chi_{k-1}) f$$

is orthogonal to $f_1, \ldots, f_{k-2}$. The function g is continuous and piecewise differentiable, and satisfies boundary conditions on each D_i, so that

$$\int\int_{D_i} |\text{grad } g|^2 \, d\mu = \lambda \int\int_{D_i} g^2 \, d\mu,$$

and when we add up, the same equality holds on D. Since g is orthogonal to $f_1, \ldots, f_{k-2}$, its Fourier expansion begins with f_{k-1}, and

$$\int\int_D |\text{grad } g|^2 \, d\mu \geqslant \lambda_{k-1} \int\int_D g^2 \, d\mu.$$

Equality holds here only if g is a linear combination of the eigenfunctions with eigenvalue λ_{k-1}, of which there are finitely many. Such a linear combination is itself an eigenfunction. Since $g \equiv 0$ on D_k, g is not an eigenfunction, equality can not hold and $\lambda > \lambda_{k-1}$.

COROLLARY. *When the boundary conditions are Neumann or periodic, the first non-trivial eigenfunction f_2 has two minimal nodal domains on D.*

The lower bounds for eigenvalues which follow are based on considering the vector $\text{grad } f^2$. We use

$$\int\int_E |\text{grad } f^2| \, d\mu = 2 \int\int_E |f| \, |\text{grad } f| \, d\mu \leqslant 2 \left\{ \int\int_E f^2 \, d\mu \int\int_E |\text{grad } f|^2 \, d\mu \right\}^{1/2}.$$

If f has a maximum or minimum other than 0 on the interior of E, the inequality is strict. If f satisfies boundary conditions on E, the right-hand side simplifies and we have

$$\int\int_E |\text{grad } f^2| \, d\mu \leqslant 2\lambda \int\int_E f^2 \, d\mu.$$

The next lemma is somewhat technical. Buser gave the first case, and the second two cases work similarly in different coordinate systems.

BUSER'S LEMMA. *Let C be an arc of a curve which cuts each straight line only finitely many times. Let E be a region bounded by C together with*

(i) *two intersecting straight lines,*

(ii) *two straight lines intersecting at infinity,*

(iii) *two non-intersecting straight lines and their common perpendicular.*

Let g be a continuously differentiable function that vanishes on C. Then

$$\int\int_E g^2 \, d\mu \leqslant 4 \int\int_E |\text{grad } g|^2 \, d\mu.$$

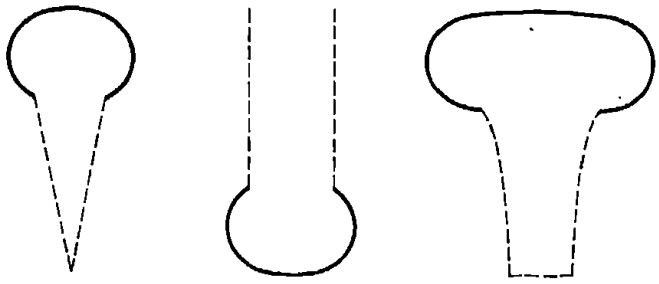

Fig. 9

If E is bounded, we may replace 4 *by* $4\text{th}^2 R/2$ *in case* (i), *where R is the furthest distance from C to the vertex, and by* $4\text{th}^2 V$ *in case* (iii), *where V is the furthest distance from C to the common perpendicular.*

Proof. In each case we take coordinates: polar in case (i), Fermi's axial coordinates (used in Section 7) in case (iii). We prove only case (ii). Although it uses familiar (x, y) coordinates, there is an extra complication, as we must consider first $E(Y)$, the set E cut off at $y = Y$. We have

$$\iint_{E(Y)} |\text{grad } y^2| \, d\mu \geqslant \iint_{E(Y)} y \left|\frac{\partial g^2}{\partial y}\right| \frac{dy\,dx}{y^2} \geqslant \int \left|\int \frac{\partial g^2}{\partial y} \frac{dy}{y}\right| dx = \int \left|\left[\frac{g^2}{y}\right] - \int g^2 \frac{dy}{y^2}\right| dx$$

$$\geqslant \iint_{E(Y)} \frac{g^2\,dx\,dy}{y^2} - \int \frac{g^2(x, Y)}{Y}\, dy,$$

since the intersections of a line $x = x_0$ with the boundary of $E(Y)$ are all on C except the point (x, y). Averaging Y over $Y_1 \leqslant Y \leqslant 2Y_1$, we get

$$\geqslant \frac{1}{\log 2} \int_{Y=Y_1}^{2Y_1} \iint_{E(Y)} \frac{g^2\,dx\,dy}{y^2} \frac{dY}{Y} - \frac{1}{\log 2} \iint_{E(2Y_1)-E(Y_1)} \frac{g^2\,dx\,dy}{y^2} \geqslant \iint_E \frac{g^2\,dx\,dy}{y^2} - o(1),$$

since g is square integrable. Expanding grad g^2 and using Cauchy's inequality as above gives the result.

TRIANGLE LEMMA. *Let D be a triangle: a domain whose boundary forms three straight lines. If g is continuously differentiable with*

$$\iint_D g \, d\mu = 0, \qquad \iint_D g^2 \, d\mu = 1,$$

then

$$\iint_D |\text{grad } g|^2 \, d\mu \geqslant 1/4.$$

Proof. By the calculus of variations, the third integral is minimised by an eigenfunction f with $\partial f/\partial n = 0$ on ∂D. Since f changes sign, there is a nodal

domain E of f on D which contains at most one corner of the triangle. E is bounded by a nodal line and possibly part of ∂D, so that f satisfies boundary conditions on E. In all configurations, case (i) or (ii) of Buser's Lemma applies, so that

$$\int\int_E f^2 \, d\mu \leqslant 4 \int\int_E |\text{grad } f|^2 \, d\mu = 4\lambda \int\int_E f^2 \, d\mu,$$

giving $\lambda \geqslant 1/4$.

BUSER'S THEOREM. *If ∂D is a polygon of r sides, with boundary conditions, then $\lambda_{r-1} \geqslant 1/4$.*

Proof. Divide D into $r-2$ triangles E_i. Some linear combination

$$g = c_1 f_1 + \ldots + c_{r-1} f_{r-1}$$

of the first $r-1$ eigenfunctions has

$$\int\int_{E_i} g \, d\mu = 0 \quad \text{for} \quad i = 1, \ldots, r-2.$$

By repeated use of the Triangle Lemma

$$\int\int_D |\text{grad } g|^2 \, d\mu \geqslant \tfrac{1}{4} \int\int_D g^2 \, d\mu,$$

so by Parseval's formula

$$\sum_1^{r-1} c_i^2 \lambda_i > \tfrac{1}{4} \sum_1^{r-1} c_i^2,$$

which is impossible if $\lambda_i < 1/4$ for $i = 1, \ldots, r-1$.

If an eigenfunction f has a closed nodal line on the upper half plane, the Triangle Lemma gives a lower bound $\geqslant 1/4$ for the eigenvalue λ. Cheeger's Theorem gives the same conclusion, with a better lower bound (except when the nodal line is a circle).

CHEEGER'S THEOREM. *Let f be a non-trivial eigenfunction. If f has a nodal domain D of finite area, then*

$$\lambda > \tfrac{1}{4}\left(1 + \frac{4\pi}{\mu(D)}\right).$$

Proof. As before

$$\int\int_D |\text{grad } f^2| \, d\mu < 2\sqrt{\lambda} \int\int_D f^2 \, d\mu = 2\sqrt{\lambda},$$

since we can suppose f to be normalised. Take curvilinear coordinates s, t, where $t = f^2$ and s is the distance along the curves $C(t)$ on which t is

constant. These coordinates are orthogonal, with

$$ds\,dt = |\text{grad}\, f^2|\, d\mu,$$

so that

$$2\sqrt{\lambda} > \int\int_D ds\,dt = \int l(t)\,dt,$$

where $l(t)$ is the length of $C(t)$. The contour lines $C(t)$ form a finite union of simple closed curves (since f is an eigenfunction and infinitely differentiable), whose interior has area $\mu(t)$. Since the circle has the greatest area for given perimeter, even in hyperbolic space,

$$\{\mu(t)+2\pi\}^2 \leqslant l^2(t)+4\pi^2,$$

$$l^2(t) \geqslant \mu^2(t)\left(1+\frac{4\pi}{\mu(t)}\right) \geqslant \mu^2(t)\left(1+\frac{4\pi}{\mu(0)}\right)$$

and

$$2\sqrt{\lambda} > \int l(t)\,dt > \left(1+\frac{4\pi}{\mu(0)}\right)^{1/2} \int \mu(t)\,dt = \left(1+\frac{4\pi}{\mu(0)}\right)^{1/2} \int\int t\,d\mu$$

on integration by parts, giving the result.

In Buser's theorem the polygon can have vertices (but not sides) at infinity, and in Cheeger's Theorem the nodal line can pass through discrete points at infinity, so long as the area remains finite. I can extend Cheeger's Theorem to one type of nodal domain with infinite area, one which is periodic in one direction only.

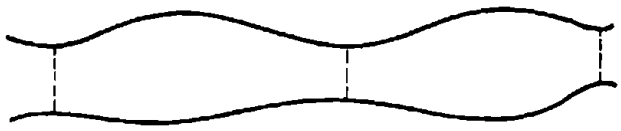

Fig. 10

When one considers n segments, there is an extra term from the periodic part of the boundary, but it remains bounded as n tends to infinity. This argument fails if the nodal domain branches with valency 3 or more.

Background reading

BANDLE
BUSER
HUXLEY

5. The group $\Gamma^0(13)$

As exceptional eigenvalues are like exceptional real zeros of Dirichlet L-functions, checking individual cases is of interest. The group of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $13 \mid b$ is called $\Gamma^0(13)$. The smallest power of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ in $\Gamma^0(13)$ is $\begin{pmatrix} 1 & 13 \\ 0 & 1 \end{pmatrix}$, so the copies of the fundamental triangle of $\Gamma(1)$ numbered -6 to 6 in the diagram are distinct under $\Gamma^0(13)$. The index is 14, so we need one more triangle to complete a fundamental domain; it has been labelled ∞.

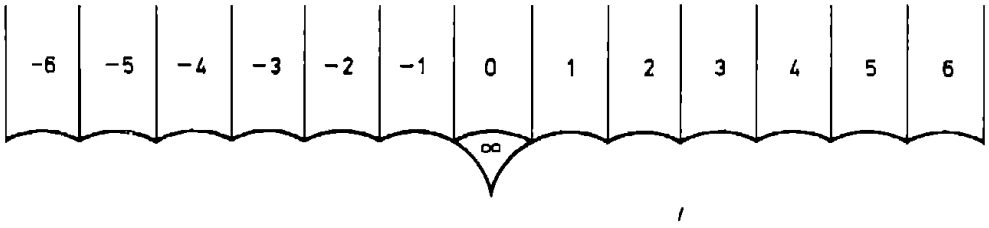

Fig. 11

Here triangle ∞ is opposite cusp ∞. Since $\begin{pmatrix} 4 & -13 \\ 1 & -3 \end{pmatrix}$ sends 3 to ∞ and ∞ to 4, region 3 is joined to region 4, and similarly $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 2 & -13 \\ 1 & -6 \end{pmatrix}$ and $\begin{pmatrix} 5 & -26 \\ 1 & -5 \end{pmatrix}$ make region 1 join region -1, region 2 join region 6 and region 5 join region 5 as shown in Figure 12.

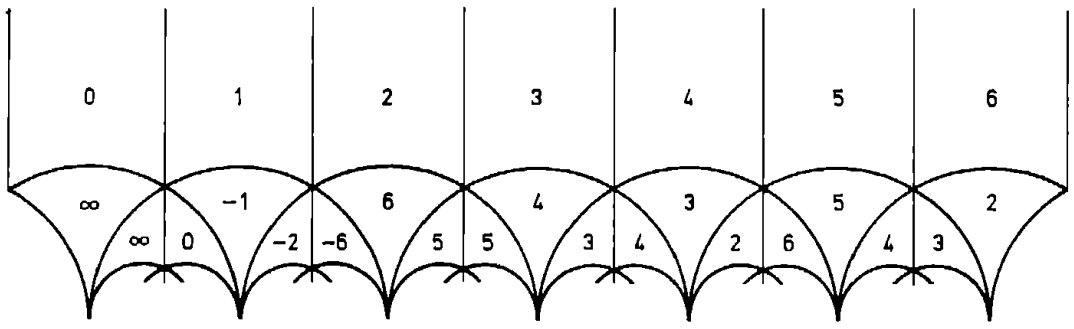

Fig. 12

The matrices $\begin{pmatrix} 4 & -13 \\ 1 & -3 \end{pmatrix}$ and $\begin{pmatrix} 5 & -26 \\ 1 & -5 \end{pmatrix}$ are rotations, the first a rotation of $2\pi/3$ about a point K, $\varrho+4$, and the second a rotation of π about a point M, $5+i$. Thus K is a branch point of order 3, and M a branch point of order 2.

The branch points can occur because $13 \equiv 1 \pmod 3$ and also (mod 4). There are two other centres of symmetry which we discover as follows. Firstly $x = 0$ is an obvious line of symmetry. Less trivially $\begin{pmatrix} 0 & -13 \\ 1 & 0 \end{pmatrix}$, representing a rotation of π about a point $A = i\sqrt{13}$, corresponds to the Hecke operator W_{13} and gives a symmetry of the fundamental domain. Conjugating by this matrix gives an outer automorphism of $\Gamma^0(13)$ sending

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \begin{pmatrix} -d & 13c \\ b/13 & -a \end{pmatrix}.$$

Since A is on $x = 0$, there is fourfold symmetry at A, and another symmetry axis $|z| = \sqrt{13}$ intersects $x = 0$ there at right angles. It passes through K and its mirror image the branch point L, $\varrho - 3$.

We now have sixfold symmetry at K, so that there are two more symmetry axes through K at angles of $2\pi/3$ to AK. One of these turns out to be KM, making M a centre of fourfold symmetry. The perpendicular to KM at M must also be a symmetry axis. It cuts $x = 13/2$ at right angles at B, $(13 + i\sqrt{13})/2$. Since $x = 13/2$ is also an obvious symmetry axis, B is also a centre of fourfold symmetry. We can now re-draw the fundamental domain.

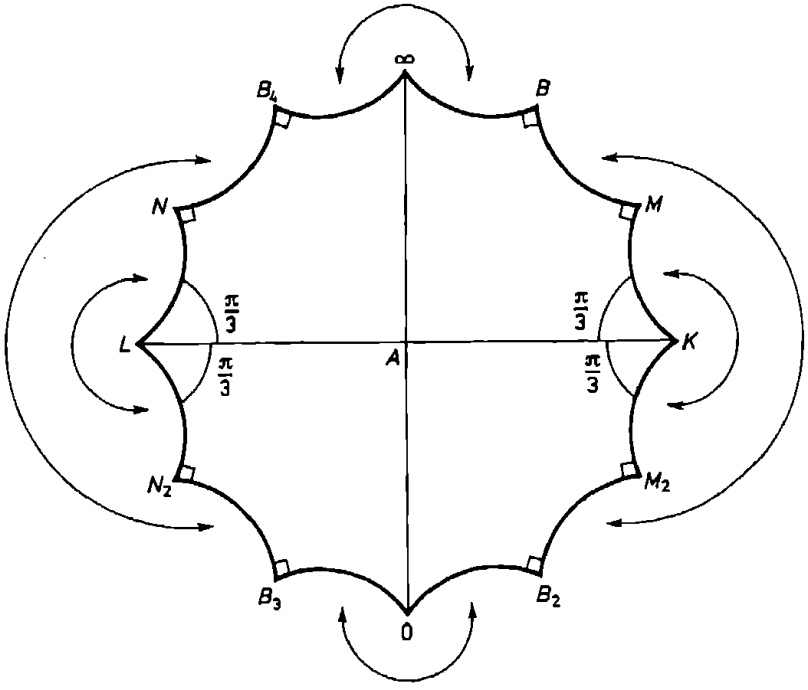

Fig. 13

The other corners are M_2, $(5+i)/2$; B_2, $(13+i\sqrt{13})/28$; B_3, $(-13+i\sqrt{13})/28$; N_2, $(-5+i)/2$; L, $\varrho+3$; N, $-5+i$ and B_4, $(-13+i\sqrt{13})/12$, all centres of symmetry. We see what is called the pin-cushion distortion in optics. We can look at the hyperbolic plane only through a wide angle lens, which compresses the whole plane into a circle or a half-plane.

The arrows indicate how the sides are identified. Since the arrows do not cross, the topological genus is 0, which is allowed when there are branch points. The subgroup $\Gamma^0(78)$ of index 12 in $\Gamma^0(13)$ has a fundamental domain twelve times as large, with no branch points, but high genus.

The fundamental domain has all these symmetry lines, but a Maass wave form can be either symmetric or skew-symmetric across symmetry lines. There are four cases.

(i) f skew symmetric across 0 to ∞ and $|z| = \sqrt{13}$. The lines 0 to ∞, LAK and ∞B are nodal. By rotation about K, KM is also a nodal line. Finally BM is also a nodal line. For if P is a point on BM, P_2 the corresponding point on $B_2 M_2$, $f(P) = f(P_2)$ since the boundary conditions join BM to $B_2 M_2$. But $f(P) = -f(P_2)$ by reflection in LAK, so $f(P) = 0$. Since f is zero on the sides of the pentagon $\infty AKMB$, whose area is $7\pi/3$, we have $\lambda > 19/28$ by Cheeger's theorem.

(ii) f skew symmetric across 0 to ∞ (and hence also across ∞ to b), but symmetric across $|z| = \sqrt{13}$. This case is best put aside.

(iii) f symmetric across 0 to ∞ and across $|z| = \sqrt{13}$.

(iv) f symmetric across 0 to ∞ and skew symmetric across $|z| = \sqrt{13}$.

We can treat cases (iii) and (iv) together. Since f is zero at the cusp $i\infty$, a nodal line passes through ∞, and we have at least one nodal line in each quadrant.

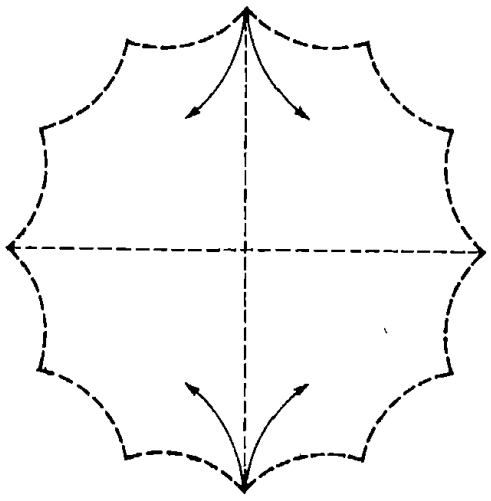

Fig. 14

The nodal line that enters the pentagon $\infty AKMB$ must cut one of the sides again, and we have five cases. If it cuts ∞A,

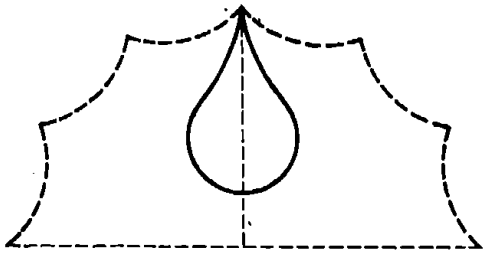

Fig. 15

there is a nodal domain of area at most $7\pi/3$, so $\lambda > 19/28$. If it cuts AK,

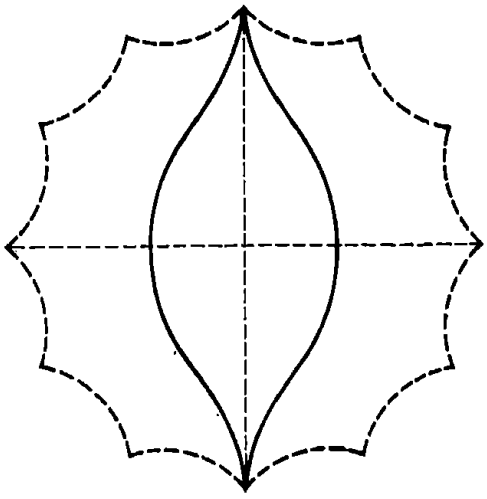

Fig. 16

there is a nodal domain of area at most $14\pi/3$, and $\lambda > 13/28$. We now redraw the fundamental domain with B at the centre.

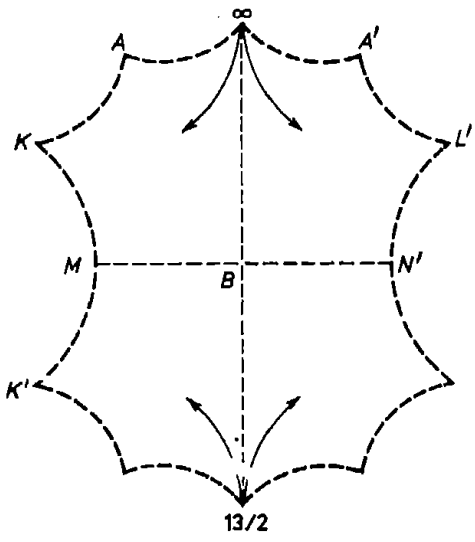

Fig. 17

We now see as before that if the nodal line cuts ∞B there is a nodal domain of area at most $7\pi/3$, and if it cuts BM there is a nodal domain of area at most $14\pi/3$.

Finally, if the nodal line crosses KM we have the diagram.

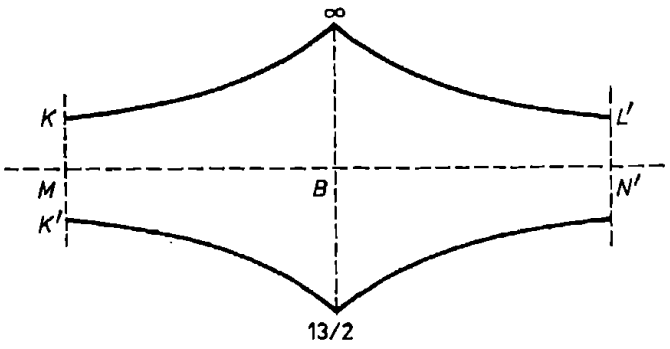

Fig. 18

Since M and N' are centres of four fold rotational symmetry, and KM, $L'N'$ are lines of reflectional symmetry, repeated reflection in KM and $L'N'$ produces a long thin nodal domain, repeating at a distance $4BM$, a so called "worm". By the extension of Cheeger's Theorem, $\lambda > 1/4$. In case (iv) we also have BM as a nodal line; the conclusion remains merely $\lambda > 1/4$.

To discuss case (ii) we draw a picture with K at the centre. This is not a fundamental domain, but three right-hand halves of a fundamental domain. This picture is consistent with $f > 0$ on the right-hand half of the fundamental domain, and $f < 0$ on the left-hand half. The copies of KMK' form a tree with valency three, so the nodal domain branches repeatedly.

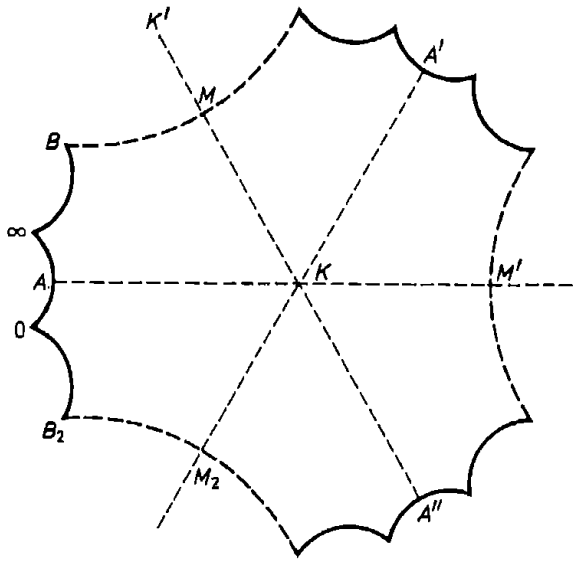

Fig. 19

This is the first case in which combinatoric geometry fails, and one has to calculate. However, there cannot be two exceptional eigenfunctions, since they must both have this pattern of nodal lines, and since f_2 is orthogonal to f_1, f_2 must change sign on the domain shown. If we put in another nodal line, there must be a bounded nodal domain or a worm, and so $\lambda_2 \geqslant 1/4$.

Background reading
ATKIN and LEHNER
RANKIN

6. Numerical bounds for the first eigenvalue

Upper bounds for eigenvalues are easier to find than lower ones. Construct any function that satisfies the boundary conditions, and calculate the integrals of g^2 and $|\text{grad } g|^2$. For instance, if you know where the nodal lines are, take g to be the distance from the nearest nodal line, so that grad g is an unit vector. If E is the nodal domain

$$\mu(E) = \int_E \int |\text{grad } g|^2 \, d\mu \geqslant \lambda \int_E \int g^2 \, d\mu,$$

so that

$$1/\lambda > d^2,$$

where d is the root mean square distance from a nodal line. We use this in the next section. The calculation could be made in the outstanding case on $\Gamma^0(13)$. I would expect it to get within a factor 2 of the true eigenvalue.

Lower bounds depend on finding a point Q where $f(Q) = 0$, drawing a curve from P to Q and using

$$f(P) = \int_Q^P \text{grad } f \, ds,$$

so that for any positive function $h(s)$

$$f^2(P) < \int_Q^P \frac{|\text{grad } f|^2}{h(s)} \, ds \int_Q^P h(s) \, ds.$$

We use a weight function as we must integrate P over the fundamental domain, and the same point Q is used for many points P. We want $h(s)$ large near Q, and small away from Q. The weight $h(s)$ should be as close to the component of grad f as possible, so it is often best to choose it as the derivative of an eigenfunction which vanishes at Q, but may not satisfy all the boundary conditions. This is quite feasible when the eigenfunction on D is skew-symmetric at the cusps. On $\Gamma^0(N)$ for $N = 13, 17, 21, 22$ and 28 the

method based on Cheeger's theorem only fails for skew-symmetric functions, and calculations by Alan Cohen in Cardiff give:

N	13	17	21	22	28
$\lambda >$	0.29...	0.28...	0.20...	0.23...	0.54...

This settles $N = 13$, 17 and 28, but not $N = 21$ or 22.

When no nodal line is known explicitly, life is much more difficult. Since our functions are cusp forms, we can try $Q = \infty$; this gives the bound $\lambda > 0$. However Hopf's rule applies even at cusps, so a nodal line passes through the cusp, and each horocycle $y =$ constant cuts some nodal line in the required point Q. In the diagram the cusp is $i\infty$.

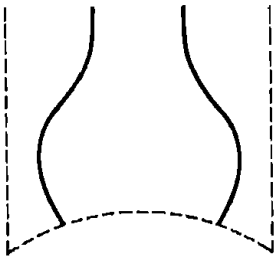

Fig. 20

We can save about five lines of argument by using the Fourier expansion, which will be of the form

$$f(x, y) \sim \sum_{n \neq 0} a_n(y)\, e\left(\frac{nx}{w}\right),$$

so that for fixed y

$$\int_{-w/2}^{w/2} |f|^2 \frac{dx}{w} = \sum_{n \neq 0} |a_n(y)|^2 \leqslant \frac{w^2}{4\pi^2} \sum_{n \neq 0} \left|\frac{2\pi i n}{w}\, a_n(y)\right|^2 = \frac{w^2}{4\pi^2} \int_{-w/2}^{w/2} \left|\frac{\partial f}{\partial x}\right|^2 \frac{dx}{w}.$$

However if P is near the bottom of the fundamental region, all points Q on the same horizontal line with $f = 0$ may be outside the fundamental domain. For a subgroup of the full modular group $\Gamma(1)$ a trick due to Roelcke and Vignéras does this quite efficiently. Draw the fundamental domain as an union of copies D_i of the triangular fundamental domain for the full modular group. Let E be the lune (region bounded by circular arcs) between $|z| = 1$ and $y = \sqrt{3}/2$, and let E_i be the image of it corresponding to D_i.

The idea is that

$$\int\int_{D_i} f^2\, d\mu \leqslant \int\int_{D_i + E_i} f^2\, d\mu.$$

There will be a certain number of triangles D_l corresponding to each distinct cusp in the fundamental region. Experts call this number the width: it is the w in the Fourier expansion above. We integrate $f^2\,d\mu$ over these w copies of

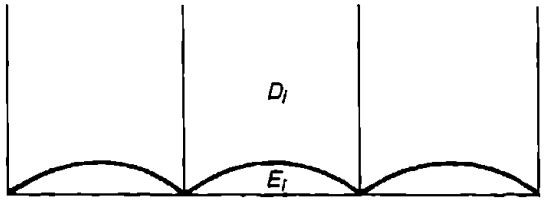

Fig. 21

$D_l \cup E_l$ using the inequality above on the integration over x, taking out $1/y^2$ at its maximum value 4/3, and w^2 at its maximum value N^2, where N is the level of the subgroup, the lowest common multiple of the widths at the cusps. Here comes the cunning part. Under the rules for joining up the fundamental domain, each E_l is a subset of some $D_{\sigma l}$, where σ is a permutation (in fact an involution) of the cosets of Γ in $\Gamma(1)$. So the integral over all the E_l is less than or equal to the integral over all the D_l, and we have

$$\int\int_D f^2\,d\mu \leqslant \frac{N^2}{4\pi^2}\,\frac{4}{3}\,2\int\int_D |\operatorname{grad} f|^2\,\frac{dx\,dy}{y^2} = \frac{2N^2}{3\pi^2}\,\lambda\int\int_D f^2\,d\mu,$$

since

$$\left|y\,\frac{\partial f}{\partial x}\right|^2 + \left|y\,\frac{\partial f}{\partial y}\right|^2 = |\operatorname{grad} f|^2$$

in the hyperbolic metric. We deduce that

$$\lambda \geqslant 3\pi^2/2N^2, \quad \lambda > 1/4 \quad \text{for} \quad N \leqslant 7.$$

When I first understood this argument, I thought it must be very wasteful; now I am not so sure. The other obvious thing to try is:

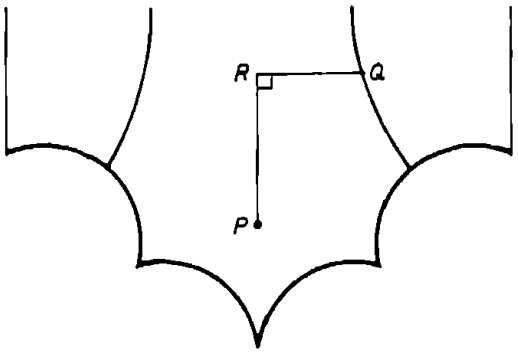

Fig. 22

Some notation; let a be the minimum of y on the domain in the picture, P be (x, y), R be (x, u). Then

$$f(x, y) = f(x, u) - \int_y^u \frac{\partial f}{\partial y}(x, v)\, dv,$$

$$f^2(x, y) \leqslant \left(1+\frac{1}{k}\right) f^2(x, u) + (1+k)\left\{\int_y^u \frac{\partial f}{\partial y}(x, v)\, dv\right\}^2$$

and the usual argument takes

$$\left\{\int_y^u \frac{\partial f}{\partial y}(x, v)\, dv\right\}^2 \leqslant \int_y^u h(v)\, dv \int_y^u \left|\frac{\partial f}{\partial y}(x, v)\right|^2 \frac{dv}{h(v)}.$$

What do we choose here? Not merely $h(y)$, but also $k(y)$ and $u(y)$ are functions of y. Their best choice depends on a, which should really be considered as a function of x. There will also be a best weight function on RQ. To optimise them all is a nasty problem in the calculus of variations, non linear because of the presence of $k(y)$. If you let $k(y)$ be constant, at least it is a linear problem, but then you lose most of the saving.

To make life easier

$$h(v) = \begin{cases} v-a & \text{for} \quad a \leqslant v \leqslant 2a, \\ a & \text{for} \quad v \geqslant 2a. \end{cases}$$

$$k(y) = \frac{\frac{7}{8}a^2}{y^2 - \frac{7}{8}a^2}, \qquad 1+\frac{1}{k} = \frac{8y^2}{7a^2}.$$

The other function $u(y)$ will be chosen shortly. We want

$$\int_a^{2a} \frac{f^2(x, y)}{y^2}\, dy \leqslant \int_a^{2a} \frac{8y^2}{7a^2} f^2(x, u) \frac{dy}{y^2} +$$

$$+ \int_a^{2a} \frac{y^2}{y^2 - \frac{7}{8}a^2}\left(a(u-2a) + \tfrac{1}{2}a^2 - \tfrac{1}{2}(y-a)^2\right) \int_y^u \left|\frac{\partial f}{\partial y}(x, v)\right|^2 \frac{dv}{v-a}\frac{dy}{y^2}.$$

In the second integral we interchange orders, making it

$$\int_{v=a}^{2a} \left|\frac{\partial f}{\partial y}(x, v)\right|^2 \int_{y=a}^{v} \frac{a(u-\frac{3}{2}a) - \frac{1}{2}(y-a)^2}{y^2 - \frac{7}{8}a^2}\, dy \frac{dv}{v-a} +$$

$$+\int_{v=2a}^{\infty}\left|\frac{\partial f}{\partial y}(x,v)\right|^2\int_{y=a}^{u^{-1}(v)}\frac{a(u-\frac{3}{2}a)-\frac{1}{2}(y-a)^2}{y^2-\frac{7}{8}a^2}\,dy\,\frac{dv}{v-a}.$$

The only thing to do here is to choose the integrand of y to be constant,

$$\frac{a(u-\frac{3}{2}a)-\frac{1}{2}(y-a)^2}{y^2-\frac{7}{8}a^2}=4,\qquad u=\tfrac{3}{2}a+\frac{(y-a)^2}{2a}+\frac{4}{a}(y^2-\tfrac{7}{8}a^2)$$

and the complicated term is

$$<4\int_{v=a}^{\infty}\left|\frac{\partial f}{\partial y}(x,v)\right|^2 dv.$$

The simple term now becomes complicated, from

$$\frac{8}{7a^2}\int_{y=a}^{2a}f^2(x,u)\,dy=\frac{8}{7a^2}\int_{u=2a}^{29a/2}f^2(x,u)\,\frac{du}{u}\leqslant\frac{8}{7a^2}\int_{u=2a}^{\infty}\frac{f^2(x,u)\,du}{8}.$$

Finally

$$\int_{2a}^{\infty}f^2(x,y)\,\frac{dy}{y^2}\leqslant\frac{1}{4a^2}\int_{2a}^{\infty}f^2(x,y)\,dy.$$

Recall that a is really a function of x: now forget it again. For simplicity we treat it as constant, $\sqrt{3}/2$ in fact. So

$$\int_{-w/2}^{w/2}\int_{2a}^{\infty}\frac{11}{28a^2}f^2\,dy\,dx\leqslant\frac{w^2}{4\pi^2}\int_{-w/2}^{w/2}\int_{2a}^{\infty}\frac{11}{28a^2}\left(\frac{\partial f}{\partial x}\right)^2 dy\,dx.$$

Now

$$|\operatorname{grad} f|^2\,d\mu=\left(\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2\right)dx\,dy,$$

so we have

$$\int_{-w/2}^{w/2}\int_{a(x)}^{\infty}f^2\,d\mu<\max\left(4,\frac{11w^2}{112\pi^2a^2}\right)\int\int|\operatorname{grad} f|^2\,d\mu.$$

So if the level N has

$$N \leqslant \pi a \sqrt{\frac{448}{11}} = 17.3\ldots,$$

we get the inequality with < 4, which gives $\lambda > 1/4$.

THEOREM. *There are no exceptional cusp form eigenvalues on any subgroup of level* $\leqslant 17$. *So there are no exceptional eigenvalues on* $\Gamma^0(N)$ *when* $N \mid m^2$ *for some integer* $m \leqslant 17$.

Note 1. This theorem covers all the cases I can do by other means.

Note 2. Finite level does not imply finite index!

Note 3. I can get $N \leqslant 18.1$ from more complicated choices of the weight functions. Machine calculations will be the next step.

Background reading

BANDLE
BUSER
HUXLEY

7. Exceptional eigenvalues in genus zero

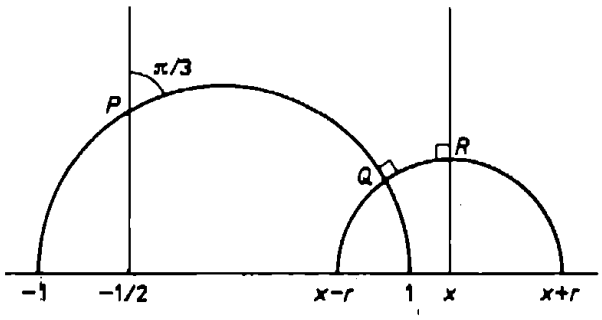

Fig. 23

In this section we give explicit examples of groups for which the first eigenvalue is less than 1/4. By varying the construction we can have the eigenfunction zero at the cusp ∞ (a cusp form) or tending slowly to infinity (a Maass wave form in the Eisenstein system). Our construction resembles Buser's Y-pieces, so that there will be subgroups for which there are arbitrarily many exceptional eigenvalues.

We start with the unit circle and the line $x = -1/2$, intersecting in P, $\exp(2\pi i/3)$. For a variable $x > 1$ we let QR be the common perpendicular from $(-1, 1)$ to (x, ∞), represented by the circle radius r, where $x^2 - r^2 = 1$.

The distance QR, d is given by

$$\text{th}^2 \frac{d}{2} = \frac{x-1}{x+1}, \qquad x = \text{ch } d, \qquad r = \text{sh } d.$$

For the distance PQ, D, we have the more complicated formula

$$\text{ch } D = \frac{|x+r-\varrho|\,|x-r-\varrho|}{2r \text{ Im } \varrho} = \frac{2x+1}{r\sqrt{3}}, \qquad \text{sh } D = \frac{x+2}{r\sqrt{3}}.$$

Reflecting the quadrilateral ∞PQR in its sides gives a tiling of the hyperbolic plane, with sixfold symmetry at P, four fold symmetry at Q and R. The area of the quadrilateral is $2\pi/3$.

First we consider eigenfunctions symmetric in ∞P, PQ and ∞R, but skew-symmetric across QR. Their symmetry group has a fundamental region consisting of four copies of the quadrilateral. It has two cusps, two branch points of order three, and genus zero.

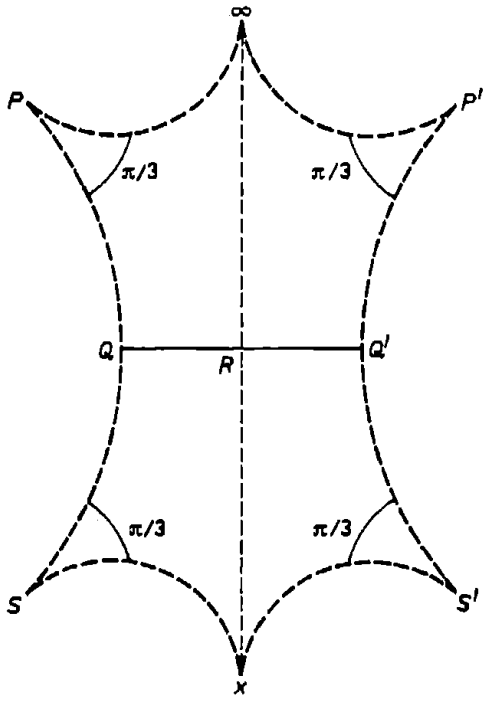

Fig. 24

To estimate the first eigenvalue, we take the Fermi hypercyclic coordinates u, v, where u is measured along QR, v is the distance perpendicular to QR. The differential of area in these coordinates can be calculated to be

$$d\mu = \text{ch } v \, du \, dv.$$

An eigenfunction with this symmetry satisfies boundary conditions on the

quadrilateral ∞PQR. Let E be the inside of this quadrilateral. Using the coordinate v as a test function, we have

$$\lambda_1 \int\int_E v^2 \, d\mu \leqslant \int\int_E |\text{grad } v|^2 \, d\mu = \int\int_E d\mu = 2\pi/3.$$

The coordinate v is not constant along ∞P. Its least value is the shortest distance c from ∞P to QR, given by

$$\text{th}^2 \frac{c}{2} = \frac{x+1/2-r}{x+1/2+r}, \qquad \text{ch } c = \frac{x+1/2}{r}.$$

Let E_1 be the set of points within E with $v \leqslant c$. Then

$$\int\int_E v^2 \, d\mu \geqslant \int\int_{E_1} v^2 \, d\mu = \int_0^d \int_0^c v^2 \text{ ch } v \, dv \, du$$

$$= d\big((c^2+2) \text{ sh } c - 2c \text{ ch } c\big) > d(c-1)^2 \text{ ch } c$$

$$= \frac{(c-1)^2 d(2x+1)}{2r} = \frac{(c-1)^2 d(2 \text{ ch } d+1)}{2 \text{ sh } d} \geqslant \tfrac{3}{2}(c-1)^2.$$

We deduce that

$$\lambda_1 < \frac{4\pi}{9(c-1)^2},$$

which is less than 1/4 for c sufficiently large, and thus for x sufficiently close to 1. We find that for $r = (x^2-1)^{1/2} < 1/20$ there must be an exceptional eigenvalue. By Courant's nodal line theorem, the first eigenfunction is zero only on the line QR and its images under the symmetry group, so it is not zero at ∞. We can also see this from Cheeger's Theorem, by drawing nodal lines through the cusps and considering all configurations, and showing that they lead to $\lambda > 1/4$. The Eisenstein series for this group will not be an Epstein zeta function, but will have a pole at β_1, where $\beta_1(1-\beta_1) = \lambda_1$.

Next we consider eigenfunctions symmetric across ∞P, QP and QR, but skew symmetric across ∞R. Again the fundamental domain for the symmetry group consists of four copies of E. There is one cusp and four branch points, two of order two and two of order three. The genus is zero.

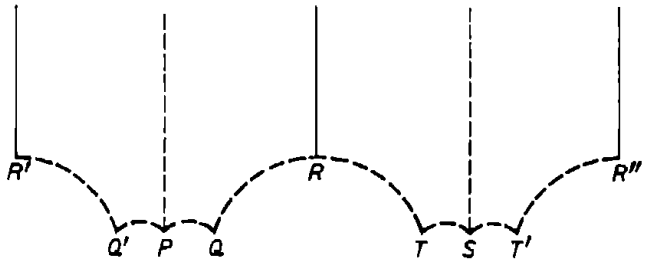

Fig. 25

We take Fermi coordinates with u measured along $R\infty$, v at right angles. As before we have

$$\lambda_1 \int\int_E v^2 \, d\mu \leqslant 2\pi/3.$$

A suitable subregion E_2 is more difficult to construct. Let the hypercycle consisting of all points distance d from ∞R cut $(-1/2, \infty)$ at M. Let N be the foot of the perpendicular from M to (x, ∞). The coordinates are

$$M: \ -\tfrac{1}{2}+\frac{i(x+\frac{1}{2})}{r}, \qquad N: \ x+i\,\frac{x(x+\frac{1}{2})}{r}.$$

The distance RN is given by

$$c = \log \frac{x(x+\frac{1}{2})}{r^2} = \log \frac{x(x+\frac{1}{2})}{x^2-1} \geqslant \frac{1}{2x}.$$

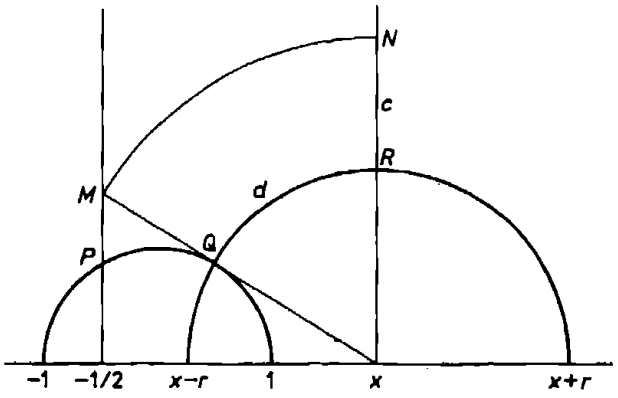

Fig. 26

We take E_2 to be bounded by QR, RN, MN and the hypercycle, with

$$\int\int_{E_2} v^2 \, d\mu = \int_0^b \int_0^d v^2 \operatorname{ch} v \, dv \, du > c(d-1)^2 \operatorname{ch} d > \tfrac{1}{2}(d-1)^2,$$

and we deduce that $\lambda_1 < 4\pi/3(d-1)^2$, which is less than $1/4$ for d sufficiently large, for instance when $x \geqslant 1982$. The first eigenfunction changes sign only on the images of (x, ∞), and is a cusp form: its mean value is zero along any horocycle centre ∞.

These eigenfunctions have symmetry groups of index two in the group generated by the rotation T_1 of π about R, the transvection T_2 at ∞ and the rotation T_3 of $2\pi/3$ about P, where

$$T_1 = \begin{pmatrix} x/r & (1-2x^2)/r \\ 1 & -x/r \end{pmatrix}, \qquad T_2 = \begin{pmatrix} 1 & 2x+1 \\ 0 & 1 \end{pmatrix}, \qquad T_3 = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Our eigenfunctions satisfy

$$f(T_i z) = \chi(T_i) f(z), \quad \chi(T_i) = \pm 1.$$

For suitable values of x the T_i are all in $\mathrm{SL}(2, Q)$.

If $m \geqslant 3964$ is an integer, there is an exceptional cusp form with

$$\chi(T_1) = -1, \quad \chi(T_2) = -1, \quad \chi(T_3) = 1,$$

where

$$T_1 = \begin{pmatrix} \dfrac{m^2+1}{m^2-1} & -\dfrac{m^4+1}{m(m^2-1)} \\ \dfrac{2m}{m^2-1} & -\dfrac{m^2+1}{m^2-1} \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & \dfrac{m^2+m+1}{m} \\ 0 & 1 \end{pmatrix}.$$

If $m \geqslant 50$ is an integer, there is an exceptional function in the Eisenstein system with

$$\chi(T_1) = -1, \quad \chi(T_2) = 1, \quad \chi(T_3) = 1,$$

and

$$T_1 = \begin{pmatrix} \dfrac{m^2+1}{2m} & -\dfrac{m^4+6m+1}{2m(m^2-1)} \\ \dfrac{m^2-1}{2m} & -\dfrac{m^2+1}{2m} \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & \dfrac{3m^2+1}{m^2-1} \\ 0 & 1 \end{pmatrix}.$$

Background reading

BANDLE
BUSER
HUXLEY

8. The spectral theorem

We say $f(z)$ is automorphic when $f(\gamma z) = f(z)$ for each γ in the discrete subgroup Γ being considered. In Bruggeman's lectures Γ was arbitrary; in this account we take $\Gamma = \mathrm{SL}(2, Z)$. For $\mathrm{SL}(2, Z)$ and its congruence subgroups the Eisenstein series are explicit functions: for $\mathrm{SL}(2, Z)$ we know from Section 1

$$E(z, s) = \sum_{\langle T \rangle \gamma} (\mathrm{Im}\ \gamma z)^s = \tfrac{1}{2} \sum_{\substack{c=-\infty \\ (c,d)=1}}^{\infty} \sum_{d=-\infty}^{\infty} \frac{y^s}{|cz+d|^{2s}}$$

for $\mathrm{Re}\ s > 1$, with a meromorphic continuation for all s, with a single pole at $s = 1$, residue $3/\pi = 1/\mu(D)$, where D is a fundamental domain. The sum over $\langle T \rangle \gamma$ means a sum over the cosets in $\mathrm{PSL}(2, Z)$ of the transvection group at

∞ generated by T: $z \to z+1$. For arbitrary discrete groups the existence of the Eisenstein series is a technical problem, and it is probably easier to follow Faddeev's approach from the theory of differential equations.

When $f(z)$ is automorphic and square integrable ($f \in L^2(\Gamma \backslash H)$) and $\operatorname{Re} s > 1$

$$\int\int_D E(z, s)\overline{f(z)}\, d\mu = \int\int_D \sum_{\langle T\rangle \gamma} (\operatorname{Im} \gamma z)^s \overline{f(z)}\, d\mu$$

$$= \sum_{\langle T\rangle \gamma} \int\int_{\gamma D} y^s \overline{f(z)}\, \frac{dx\, dy}{y^2} = \int_0^\infty \int_{-1/2}^{1/2} y^s \overline{f(z)}\, \frac{dx\, dy}{y^2}$$

$$= \int_0^\infty y^s \overline{f_0(y)}\, \frac{dy}{y^2},$$

where $f_0(y)$ is the term independent of x in the Fourier series

$$f(z) \sim \sum_{-\infty}^{\infty} f_n(y)\, e(nx).$$

This remains true by analytic continuation as long as both sides converge absolutely. However to ensure convergence we work with continuous functions of compact support. Let $g(y)$ be a piecewise differentiable continuous function, non zero only on a bounded interval $a \leqslant y \leqslant b$. Let

$$G(s) = \int_a^b \frac{g'(y)}{y^s}\, \frac{ds}{s} = \int_0^\infty \frac{g(y)\, dy}{y^{s+1}}$$

be the Mellin transform of $g(y)$; it is an entire function. We require that for each fixed σ as $t \to \infty$

$$G(\sigma + it) \ll 1/t^2;$$

this certainly holds if $g(y)$ is twice differentiable. Let L be a line $\operatorname{Re} s = \sigma_0$, with $\sigma_0 > 1$. By the usual Perron's formula argument

$$\frac{1}{2\pi i} \int_L G(s)\, x^2\, ds = \int_a^b \frac{1}{2\pi i} \int_L \left(\frac{x}{y}\right)^s \frac{ds}{s}\, g'(y)\, dy = \int_a^x g'(y)\, dy = g(x).$$

We form an automorphic function by

$$H(z) = \sum_{\langle T\rangle \gamma} g(\operatorname{Im} \gamma z).$$

The sum is finite, but we may rearrange it as follows:

$$H(z) = \sum_{\langle T\rangle_\gamma} \frac{1}{2\pi i} \int_L (\operatorname{Im} \gamma z)^s G(s)\, ds = \frac{1}{2\pi i} \int_L E(z, s)\, G(s)\, ds$$

where we may interchange summation and integration by absolute convergence. On moving the contour we see that

$$H(z) = \frac{3}{\pi} G(1) + \frac{1}{2\pi i} \int_{\mathrm{Res}=1/2} G(s) E(z, s)\, ds.$$

When $f(z)$ is automorphic, the same argument used above with the Eisenstein series gives

$$(H, f) = \int_D \int H(z)\overline{f(z)}\, d\mu = \int_0^\infty g(y)\overline{f_0(y)}\, \frac{dy}{y^2}.$$

As special cases

$$(H, 1) = \int_0^\infty \frac{g(y)\, dy}{y^2} = G(1),$$

$$(H(z), E(z, s)) = \int_0^\infty g(y) \left\{ y^s + \left(\frac{\overline{E(z, s)}}{E(z, 1-s)} \right) y^{1-s} \right\} \frac{dy}{y^2}$$

$$= G(1-\bar{s}) + \frac{E(z, \bar{s})}{E(z, 1-\bar{s})}\, G(\bar{s}).$$

We now have

$$\int_{\mathrm{Res}=1/2} (H(z), E(z, s)) E(z, s)\, dt = \int_{\mathrm{Res}=1/2} \{G(s) E(z, s) + G(\bar{s}) E(z, \bar{s})\}\, dt$$

$$= 2 \int_{\mathrm{Res}=1/2} G(s) E(z, s)\, dt,$$

so that we can write

$$H(z) = \sqrt{\frac{3}{\pi}} \left(H(z), \sqrt{\frac{3}{\pi}} \right) + \frac{1}{4\pi} \int_{\mathrm{Res}=1/2} (H(z), E(z, s)) E(z, s)\, dt.$$

This implies a Parseval theorem

$$(H, H) = \left| \left(H(z), \sqrt{\frac{3}{\pi}} \right) \right|^2 + \frac{1}{4\pi} \int_{\mathrm{Res}=1/2} |(H(z), E(z, s))|^2\, dt.$$

We define the cusp space M_c to be those automorphic functions for which

$$\int\int_D |f(z)|^2 \, d\mu \quad \text{converges,}$$

$$f_0(y) = 0.$$

If $f \in M_c$, then for any $H(z)$ as constructed above $(H, f) = 0$. If $f \notin M_c$, there is a permissible function $g(y)$ with

$$\int_0^\infty g(y)\overline{f_0(y)} \, \frac{dy}{y^2} \neq 0,$$

so for the corresponding $H(z)$, $(H, f) \neq 0$. By Hilbert space theory we can write any square-integrable automorphic f as $f_E + f_c$, where $f_c \in M_c$, $(f_E, \varphi) = 0$ for any φ in M_c, and f_E can be approximated by a function $H(z)$ in the sense that for any ε, there is an $H(z)$ with

$$(f_E - H, f_E - H) < \varepsilon^2.$$

Hence for any fixed T

$$(f_E, f_E) \geqslant \left|\left(f, \sqrt{\frac{3}{\pi}}\right)\right|^2 + \frac{1}{4\pi} \int_{1/2-iT}^{1/2+iT} |(f, E(z, s))|^2 \, dt,$$

and as $T \to \infty$ the right-hand side approaches the left.

In the cusp space we shall have orthogonal eigenfunctions $f_1, f_2, \ldots$ with

$$(f_c, f_c) = \sum_1^\infty |(f_c, f_i)|^2.$$

To prove this it is sufficient to show that for any $\nu > 0$ there is an N such that any N-dimensional space of square-integrable automorphic functions contains a function f with

$$\int\int_D |\operatorname{grad} f|^2 \, d\mu \geqslant \nu \int\int_D f^2 \, d\mu.$$

We divide the fundamental region for the full modular group into triangles, of which one contains the cusp, the others are bounded (Fig. 27). We restate some calculations from Section 6.

CUSP LEMMA. *Suppose* $f(x, y)$ *satisfies for each* y

$$f(x+w, y) = f(x, y) \quad \text{and} \quad \int_{-w/2}^{w/2} f(x, y) \, dx = 0.$$

Let E be a domain bounded by $x = \pm w/2$ and a curve $y = a(x)$. Choose $b \geqslant \max a(x)$, $c \leqslant a = \min a(x)$. Then

$$\int\int_E |f^2|\, d\mu \leqslant K \int\int_E |\operatorname{grad} f|^2\, d\mu,$$

where

$$K = \max\left(\frac{(b-a)^2}{2(a^2-c^2)}, \frac{4\pi^2}{w^2}\left(\frac{1}{b^2}+\frac{d^2-c^2}{a(b-a)c^2}\right)\right).$$

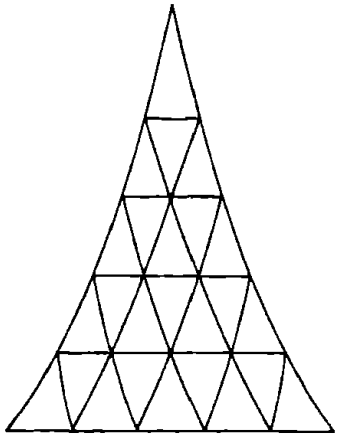

Fig. 27

We take E to be the triangle containing the cusp, so that $w = 1$, $a(x) = (a_2^2 - x^2)^{1/2}$, $a_1 = (a_2^2 - 1/4)^{1/2}$. If a_2 is large enough, we may choose b and c so that $K \leqslant 1/\nu$.

Having now chosen a_2, we divide the rest of the fundamental domain into triangles $E_2, \ldots, E_N$ of side $\leqslant 2 \operatorname{arg\,th}(1/2\sqrt{\nu})$. Any N-dimensional subspace of the cusp space M_c contains a function f with

$$\int\int_{E_i} f\, d\mu = 0 \quad \text{for} \quad i = 2, \ldots, N,$$

and by Buser's lemma and the Triangle lemma

$$\int\int_{E_i} |f|^2\, d\mu \leqslant \frac{1}{\nu} \int\int_{E_i} |\operatorname{grad} f|^2\, d\mu,$$

so that summing over the N triangles gives the inequality required.

From this an expert in Hilbert space theory (which I am not) concludes that the Laplacian restricted to the cusp form space corresponds to the inverse of a compact self-adjoint operator. Hence the cusp form space can be spanned by orthogonal square-integrable eigenfunctions, the Maass wave forms. The words 'cusp form' are usually reserved for Maass wave forms in

the cusp form space M_c. We number the orthogonal Maass wave forms $f_1, f_2, \ldots$, and normalise them by

$$(f_i, f_j) = \delta_{ij}.$$

If we write $f_0 = \sqrt{3/\pi}$, the Parseval formula is

$$(f,f) = (f_E, f_E) + (f_c, f_c) = \sum_0^\infty |(f, f_i)|^2 + \frac{1}{4\pi} \int_{-\infty}^{\infty} \left|\left(f, E\left(z, \tfrac{1}{2}+it\right)\right)\right|^2 dt.$$

If f and h are two square integrable automorphic functions, we deduce that

$$(f, h) = \sum_{\rho_i}^{\infty} (f, f_i)(f_i, h) + \frac{1}{4\pi} \int_{\mathrm{Re}\, s = 1/2} (f, E(z, s))(E(z, s), h)\, dt.$$

Further reading
KUBOTA
SHOWALTER

9. Bessel functions

A section on Bessel functions is needed, firstly because the properties of Bessel functions which we use are rather recondite – some discovered by Kuznetsov himself – and more importantly, because the Bessel function is a continuous analogue of the Kloosterman sum. Let

$$F(x, \lambda) = \exp\left(\tfrac{1}{2} x\left(\lambda + \frac{1}{\lambda}\right)\right) = \sum_{-\infty}^{\infty} I_n(x) \lambda^n.$$

The Bessel coefficients $I_n(x)$ are means of $F(x, \lambda)$ round the unit circle. Since

$$\left(x \frac{\partial}{\partial x}\right)^2 F - \left(\lambda \frac{\partial}{\partial \lambda}\right)^2 F = x^2 F,$$

we have

$$\left(x \frac{d}{dx}\right)^2 I_n = (x^2 + n^2) I_n.$$

Similarly

$$G(x, \lambda) = \exp\left(\tfrac{1}{2} x\left(\lambda - \frac{1}{\lambda}\right)\right) = \sum_{-\infty}^{\infty} J_n(x) \lambda^n$$

gives

$$\left(x \frac{d}{dx}\right)^2 J_n = (n^2 - x^2) J_n.$$

We shall require $I_s(x)$ when s is not an integer. If

$$f(x) = \int_M F(x, \lambda) \frac{d\lambda}{\lambda^{s+1}},$$

$$\left(x \frac{\partial}{\partial x}\right)^2 f - x^2 f = \int_M \left(\lambda \frac{\partial}{\partial \lambda}\right)^2 F \frac{d\lambda}{\lambda^{s+1}} = (-s)^2 \int_M \frac{F\, d\lambda}{\lambda^{s+1}},$$

provided that the contour M is such that the integrated terms vanish when we integrate by parts. For $\operatorname{Re} x > 0$ we can take M to be a loop from $-\infty$ around the origin and back to $-\infty$, and put

$$I_s(x) = \frac{1}{2\pi i} \int_M F(x, \lambda) \frac{d\lambda}{\lambda^{s+1}};$$

for other nonzero x we rotate the loop M so that $\operatorname{Re} \lambda x \to -\infty$ as $|\lambda| \to \infty$ on M. A variable change $u = \frac{1}{2}\lambda x$ gives

$$\frac{I_s(x)}{x^s} = \frac{1}{2\pi i} \int_M \frac{1}{2^s} \exp\left(u + \frac{x^2}{4u}\right) \frac{du}{s+1};$$

the right-hand side is an entire function of x. If s is an integer, the integral along the loop M is the same as the integral round the unit circle, and we recover the previous definition. If $\operatorname{Re} s < 0$ the integral along M is the same as the integral along the negative real axis taken twice with arg u changed by 2π. We see that

$$\lim_{x \to 0} \frac{I_s(x)}{x^s} = \frac{1}{2^{s+1} \pi i} \int_{t=0}^{\infty} e^{-t} \{\operatorname{cis}(-s\pi) - \operatorname{cis} s\pi\} \frac{dt}{t^{s+1}}$$

$$= -\frac{\sin s\pi}{\pi 2^s} \cdot \underline{\lfloor -s-1} = -\frac{1}{2^s \pi} \cdot \frac{s\pi}{\underline{\lfloor -s} \, \underline{\lfloor s}} \cdot \underline{\lfloor -s-1} = \frac{1}{2^s \, \underline{\lfloor s}},$$

where we have used various properties of the factorial function. By analytic continuation the integral still represents $1/2^s \cdot \underline{\lfloor s}$ for all s. The differential equation now gives the familiar power series

$$I_s(x) = \sum_{r=0}^{\infty} \frac{(\frac{1}{2}x)^{s+2r}}{\underline{\lfloor r} \, \underline{\lfloor r+s}},$$

and similarly

$$J_s(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{\underline{\lfloor r} \, \underline{\lfloor r+s}} \left(\frac{x}{2}\right)^{2r+s}.$$

From the integral definition or from the power series various recurrences follow:

$$xJ_s' + sJ_s = xJ_{s-1}, \quad xJ_s' - sJ_s = -xJ_{s+1},$$

$$J_{s-1} + J_{s+1} = \frac{2s}{x} J_s, \quad J_{s-1} - J_{s+1} = 2J_s'$$

and one which Kuznetsov uses

$$(s-1)J_{s-1} - (s+1)J_{s+1} = 2sx \frac{d}{dx} \frac{J_s}{x}.$$

Kuznetsov also requires Neumann's inversion of the power series:

$$\left(\frac{x}{2}\right)^s = \sum_{n=0}^{\infty} (s+2n) \frac{\lfloor s+n-1}{\lfloor n} J_{s+2n}(x).$$

To see this we look at the coefficient of $(\frac{1}{2}x)^{s+2u}$ in the expansion (which is absolutely convergent). It is

$$\sum_{n+r=u} \frac{(s+2n)\lfloor s+n-1 \ (-1)^r}{\lfloor n \ \lfloor r \ \lfloor r+s+2n}$$

$$= \frac{\lfloor s-1}{\lfloor u \ \lfloor s+2u} \sum_{r+n=u} (s+2n)(-s)\ldots(-s-n+1)(s+2u)\ldots(s+2u-r+1) \times$$

$$\times \frac{\lfloor u}{\lfloor n \ \lfloor u-n},$$

and the sum over n is the value at $x = 1$ of

$$s\left(\frac{d}{dx}\right)^u x^{-s} x^{s+2u} - 2su\left(\frac{d}{dx}\right)^{u-1} x^{-s-1} x^{s+2u},$$

which is s for $u = 0$, but for $u > 0$ it is

$$s \frac{\lfloor 2u}{\lfloor u} - 2su \frac{\lfloor 2u-1}{\lfloor u} = 0.$$

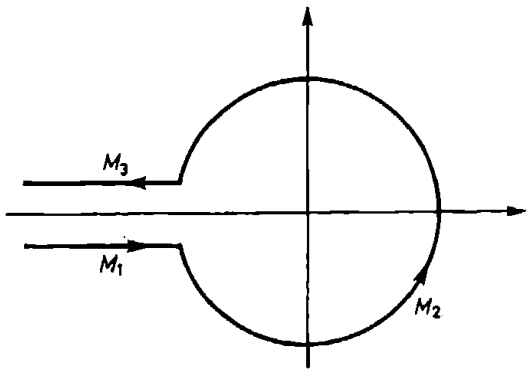

Fig. 28

Returning to the integral definition, we can take the loop as:
M_1: straight line $-\infty$ to -1,
M_2: the unit circle,
M_3: straight line -1 to $-\infty$.
On M_1 and M_3 we put $\lambda = e^t \operatorname{cis} \pm\pi$, on M_2 we put $\lambda = \operatorname{cis} \theta$. Then

$$I_s(x) = \frac{1}{\pi} \int_{t=0}^{\infty} \exp(-x \operatorname{ch} t) \sin s\pi \cdot e^{-st}\, dt + \frac{1}{2\pi} \int_0^{2\pi} \exp(x \cos\theta) \cos s\theta\, d\theta;$$

the integral with $\sin s\theta$ is zero by symmetry. The integral round the unit circle can be eliminated by subtraction:

$$I_{-s}(x) - I_s(x) = \sin s\pi \cdot \mathrm{Kh}_s(x),$$

$$\mathrm{Kh}_s(x) = \frac{2}{\pi} \int_{t=0}^{\infty} \exp(-x \operatorname{ch} t) \operatorname{ch} st\, dt,$$

which is Macdonald's integral, valid for $\operatorname{Re} x > 0$. It can be transformed as follows:

$$\mathrm{Kh}_s(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \exp(st - x \operatorname{ch} t)\, dt$$

$$= \frac{1}{\pi} \int_0^{\infty} \exp\left(-\frac{x}{2}\left(\lambda + \frac{1}{\lambda}\right)\right) \lambda^{s-1}\, d\lambda = \frac{1}{\pi} \int_0^{\infty} \frac{\lambda^{s-1}\, d\lambda}{F(x, \lambda)},$$

another integral of the same type as the one defining $I_s(x)$.

We can integrate $G(x, \lambda)$ from 0 to 1 along the real axis. This leads to Hankel's integrals

$$\mathrm{Hs}_s(x) = \frac{2}{2\pi i} \int_S G(x, \lambda) \frac{d\lambda}{\lambda^{s+1}}, \qquad \mathrm{Hi}_s(x) = \frac{2}{2\pi i} \int_I G(x, \lambda) \frac{d\lambda}{\lambda^{s+1}},$$

where the contours consist of
S_1: straight line 0 to 1,
S_2: clockwise semicircle from 1 to -1, with $\operatorname{Im} \lambda > 0$,
S_3: straight line -1 to $-\infty$;
I_1: straight line $-\infty$ to -1,
I_2: clockwise semicircle from -1 to 1, with $\operatorname{Im} \lambda < 0$,
I_3: straight line 1 to 0.

Then

$$2J_s = \mathrm{Hs}_s + \mathrm{Hi}_s,$$

and the other 'familiar' Bessel function is given by

$$2iY_s = \mathrm{Hs}_s - \mathrm{Hi}_s.$$

There are now many relations between our six functions.

$$\begin{aligned}
I_s(x \text{ cis } (\pi/2)) &= \text{cis } (s\pi/2)\, J_s(x),\\
J_s(x \text{ cis } (\pi/2)) &= \text{cis } (s\pi/2)\, I_s(x),\\
I_s(x \text{ cis } \pi) &= \text{cis } (s\pi)\, I_s(x),\\
J_s(x \text{ cis } \pi) &= \text{cis } (s\pi)\, J_s(x),\\
\mathrm{Kh}_{-s}(x) &= \mathrm{Kh}_s(x),\\
\mathrm{Hs}_s(x) &= \text{cis } (-s\pi)\, \mathrm{Hs}_{-s}(x),\\
\mathrm{Hi}_s(x) &= \text{cis } (s\pi)\, \mathrm{Hi}_{-s}(x),\\
\mathrm{Kh}_s(x \text{ cis } (\pi/2)) &= i \text{ cis } (-s\pi/2)\, \mathrm{Hi}_s(x),\\
\mathrm{Kh}_s(x \text{ cis } (-\pi/2)) &= -i \text{ cis } (s\pi/2)\, \mathrm{Hs}_s(x).
\end{aligned}$$

Hence

$$J_s(x) + J_{-s}(x) = -i \cos \frac{s\pi}{2} \{\mathrm{Kh}_s(x \text{ cis } (\pi/2)) - \mathrm{Kh}_s(x \text{ cis } (-\pi/2))\},$$

$$J_s(x) - J_{-s}(x) = \sin \frac{s\pi}{2} \{\mathrm{Kh}_s(x \text{ cis } (\pi/2)) + \mathrm{Kh}_s(x \text{ cis } (-\pi/2))\}.$$

In the Kuznetsov theory s is usually pure imaginary, and Macdonald's integral

$$\mathrm{Kh}_s(z) = \frac{2}{\pi} \int_0^\infty \exp(-z \text{ ch } t) \text{ ch } st\, dt$$

still converges on $\mathrm{Re}\, z = 0$, so by continuity the integral still represents $\mathrm{Kh}_s(z)$, and we have

$$\mathrm{Kh}_{2iu}(iy) - \mathrm{Kh}_{2iu}(-iy) = \frac{2}{\pi} \int_0^\infty 2i \sin(y \text{ ch } t) \cos 2tu\, dt,$$

$$\mathrm{Kh}_{2iu}(iy) + \mathrm{Kh}_{2iu}(-iy) = \frac{2}{\pi} \int_0^\infty 2 \cos(y \text{ ch } t) \cos 2tu\, dt,$$

and

$$J_{2iu}(y)+J_{-2iu}(y)=\frac{4\operatorname{ch}\pi u}{\pi}\int_0^\infty \sin(y\operatorname{ch}t)\cos 2tu\,dt,$$

$$J_{2iu}(y)-J_{-2iu}(y)=\frac{4i\operatorname{sh}\pi u}{\pi}\int_0^\infty \cos(y\operatorname{ch}t)\cos 2tu\,dt.$$

These integrals converge at infinity although the integrand does not tend to zero, since integration by parts yields

$$\int_T^\infty \sin(y\operatorname{ch}t)\cos 2tu\,dt$$

$$=\left[\frac{-\cos(y\operatorname{ch}t)\cos 2tu}{y\operatorname{sh}t}\right]_T^\infty-\frac{2u}{y}\int_T^\infty\frac{\cos(y\operatorname{ch}t)\sin 2tu\,dt}{\operatorname{sh}t}$$

$$\ll(1+|u|)\,y^{-1}e^{-T}.$$

At times we will wish to know the orders of magnitude of Bessel functions. When $x^2=o(|s|)$, the first term dominates in the power series for J_s and I_s. For large x we estimate by the saddle point method. The integrals involving $F(x,\lambda)$ and $G(x,\lambda)$ are suitable: $F(x,\lambda)$ has stationary phase at $\lambda=\pm1$, $G(x,\lambda)$ at $\lambda=\pm i$. For x real and large compared with s, the results are

$$I_s(x)\sim\frac{\exp x}{\sqrt{(2\pi x)}},$$

$$J_s(x)\sim\cos\left(x-\frac{s\pi}{2}-\frac{\pi}{4}\right)\sqrt{(2/\pi x)},$$

$$\mathrm{Kh}_s(x)\sim\exp(-x)\sqrt{(2/\pi x)},$$

$$\mathrm{Hs}_s(x)\sim\operatorname{cis}\left(x-\frac{s\pi}{2}-\frac{\pi}{4}\right)\sqrt{(2/\pi x)},$$

$$\mathrm{Hi}_s(x)\sim\operatorname{cis}\left(-x+\frac{s\pi}{2}+\frac{\pi}{4}\right)\sqrt{(2/\pi x)},$$

$$Y_s(x)\sim\sin\left(x-\frac{s\pi}{2}-\frac{\pi}{4}\right)\sqrt{(2/\pi x)}.$$

These bounds actually hold uniformly in regions of the type $\mathrm{Re}\, x \geqslant R$, $|\arg x| \leqslant (\pi/2)-\delta$. We shall use them only for real $x \to +\infty$ to verify that certain integrals converge.

We are taking the basic properties of the factorial function as well known (the functional equation, Euler's integral and Legendre's duplication formula). It is surprising how factorials appear in integrals with no exponential factor.

BETA-FUNCTION LEMMA. *For* $\mathrm{Re}\, \alpha > -1$, $\mathrm{Re}\, \beta > -1$

$$\int_0^\infty \frac{u^\alpha}{(1+u)^{\alpha+\beta+2}}\, du = \frac{\underline{|\alpha}\ \underline{|\beta}}{\underline{|\alpha+\beta+1}}.$$

Proof. By Euler's integral

$$\underline{|\alpha}\ \underline{|\beta} = \int_{x=0}^\infty \int_{y=0}^\infty e^{-(x+y)} x^\alpha y^\beta\, dy\, dx = \int_{z=0}^\infty e^{-z} \int_{x=0}^z x^\alpha (z-x)^\beta\, dx\, dz.$$

Now put $u = x/y = x/(z-x)$. We find

$$\underline{|\alpha}\ \underline{|\beta} = \int_{z=0}^\infty \int_{u=0}^\infty e^{-z} z^{\alpha+\beta+1} \frac{u^\alpha}{(1+u)^{\alpha+\beta+2}}\, du\, dz$$

which gives the result.

We can now prove Mellin transform theorems involving Bessel functions.

THEOREM. *For* $\mathrm{Re}\, \alpha - |\mathrm{Re}\, \beta| > -1$

$$\int_0^\infty u^\alpha \mathrm{Kh}_\beta(u)\, du = \frac{2^\alpha}{\pi} \frac{\underline{|\alpha+\beta-1}}{2} \frac{\underline{|\alpha-\beta-1}}{2},$$

$$\int_0^\infty u^\alpha e^{-u} \mathrm{Kh}_\beta(u)\, du = \frac{\underline{|\alpha+\beta}\ \underline{|\alpha-\beta}}{2^\alpha \sqrt{\pi}\ \underline{|\alpha+1/2}}.$$

Proof. Using Macdonald's integral we have

$$\int_0^\infty u^\alpha \mathrm{Kh}_\beta(u)\, du = \frac{2}{\pi} \int_{u=0}^\infty \int_{t=0}^\infty u^\alpha e^{-u \mathrm{ch}\, t} \mathrm{ch}\, \beta t\, dt\, du$$

$$= \frac{2}{\pi} \underline{|\alpha} \int_0^\infty \frac{\mathrm{ch}\, \beta t}{(\mathrm{ch}\, t)^{\alpha+1}}\, dt = \frac{\underline{|\alpha}}{\pi} \int_{-\infty}^\infty \frac{e^{\beta t}}{(\mathrm{ch}\, t)^{\alpha+1}}\, dt;$$

the double integral converges absolutely, so we may interchange integrations. When we put $v = e^{2t}$, the integral becomes

$$\frac{\lfloor\alpha}{\pi} 2^{\alpha+1} \int_0^\infty \frac{v^{(\alpha+\beta+1)/2}}{(v+1)^{\alpha+1}} \frac{dv}{2v}.$$

Similarly

$$\begin{aligned}\int_0^\infty u^\alpha e^{-u} \mathrm{Kh}_\beta(u)\, du &= \frac{2}{\pi} \int_{u=0}^\infty u^\alpha e^{-u(1+\mathrm{ch}\, t)} \mathrm{ch}\, \beta t\, dt\, du \\ &= \frac{2\,\lfloor\alpha}{\pi} \int_{t=0}^\infty \frac{\mathrm{ch}\, \beta t}{(1+\mathrm{ch}\, t)^{\alpha+1}}\, dt = \frac{\lfloor\alpha}{\pi} \int_{-\infty}^\infty \frac{e^{\beta t}}{\{\frac{1}{2}(e^{t/2}+e^{-t/2})\}^{\alpha+1}}\, dt \\ &= \frac{\lfloor\alpha}{\pi} \int_0^\infty \frac{v^\beta (2v)^{\alpha+1}}{(v+1)^{2\alpha+2}} \frac{dv}{v},\end{aligned}$$

when we put $v = e^t$. We have simplified the answer using Legendre's duplication rule

$$\lfloor 2\alpha+1 = 2^{2\alpha+1}\, \lfloor\alpha\, \lfloor\alpha+\tfrac{1}{2}/\sqrt{\pi}.$$

We also require the inverse formulae connecting factorials with Bessel functions.

THEOREM. *Let x be real and positive, and let M be a contour from $-\infty$ to the finite plane and back to $-\infty$ such that all the poles of the integrands lie to the left of M. Then*

$$\frac{1}{2\pi i} \int_M \frac{\lfloor s-\frac{3}{2}+ia\, \lfloor s-\frac{3}{2}-ia}{x^{2s}}\, ds = -\frac{\pi}{x} \mathrm{Kh}_{2ia}(2x),$$

$$\frac{1}{2\pi i} \int_M \frac{\cos \pi s\, \lfloor s-\frac{3}{2}+ia\, \lfloor s-\frac{3}{2}-ia}{x^{2s}}\, ds = \frac{-\pi}{2x\, \mathrm{ch}\, \pi a} \{J_{2ia}(2x)+J_{-2ia}(2x)\},$$

$$\frac{1}{2\pi i} \int_M \frac{\sin \pi s\, \lfloor s-\frac{3}{2}+ia\, \lfloor s-\frac{3}{2}-ia}{x^{2s}}\, ds = \frac{\pi i}{2x\, \mathrm{sh}\, \pi a} \{J_{2ia}(2x)-J_{-2ia}(2x)\}.$$

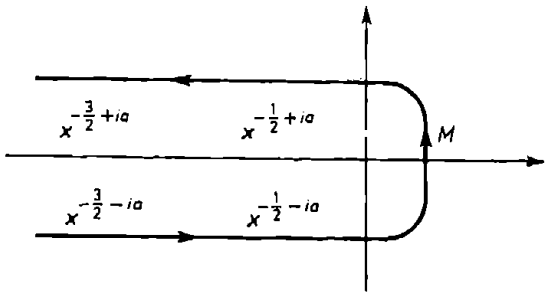

Fig. 29

Here a is not required to be real, and M may be replaced by a line L: $\operatorname{Re} s = c$, with indentations to carry the contour to the right of the poles. For convergence on L in the second and third integrals we must have $C < 0$.

Proof. If σ_0 is not the abscissa of a pole, the integral along the straight line segment $\sigma_0 - N - iT$ to $\sigma_0 - N + iT$, where T is fixed, tends to zero like $x^{2N}/(\underline{|N})^2$ when N tends to infinity through integer values. Hence the integrals are the sum of the residues at the poles. The poles of $\underline{|s-3/2+ia}$ are at $s = -n+1/2-ia$, with residue $(-1)^n/\underline{|n}$, for $n = 0, 1, 2, \ldots$ Taking the third integral as an example, the residue at this pole is

$$x^{2n-1+2ia}\frac{(-1)^n}{\underline{|n}}\,\underline{|-n-1-2ia}\,\sin\pi(-n+\tfrac{1}{2}-ia)$$

$$= \frac{x^{2n-1+2ia}}{\underline{|n}}\,\frac{\operatorname{ch}\pi a}{\underline{|n+1+2ia}}\,\frac{\pi(n+1+2ia)}{\sin\pi(n+1+2ia)}$$

$$= \frac{x^{2n-1+2ia}}{\underline{|n}}\,\frac{\operatorname{ch}\pi a}{\underline{|n+2ia}}\,\frac{\pi}{(-1)^{n+1}i\operatorname{sh}2\pi a}$$

$$= \frac{\pi i}{2x\operatorname{sh}\pi a}(-1)^n\frac{x^{2n+2ia}}{\underline{|n}\,\underline{|n+2ia}}$$

which leads to the series expansion of the Bessel function.

COROLLARY. *Writing $ia = u$ and rearranging*

$$J_{2u}(2x) = \frac{ix}{2\pi^2}\int_M \frac{\cos\pi(s-u)\,\underline{|s-\frac{3}{2}+u}\;\underline{|s-\frac{3}{2}-u}}{x^{2s}}\,ds$$

$$= \frac{x}{2\pi i}\int_M \frac{\underline{|s-\frac{3}{2}+u}}{\underline{|\frac{1}{2}+u-s}}\,\frac{ds}{x^{2s}} = \frac{1}{2\pi i}\int_M \frac{\underline{|s-1+u}}{\underline{|u-s}}\,\frac{ds}{x^{2s-1}}.$$

This integral could also have been checked directly by adding the residues.

Further integrals involving Bessel functions will be needed, but they are rather special to the Kuznetsov theory, and will be discussed in the appropriate context.

Further reading
JEFFREYS and JEFFREYS

10. Fourier coefficients of cusp forms

From the spectral theorem we know that Maass wave forms

$$f_j = \sum_1^\infty (b_n^{(j)} e(nx) + b_{-n}^{(j)} e(-nx)) A_{i\varkappa_j}(2\pi n y)$$

exist and span the cusp space. In fact the proof of the spectral theorem shows that the number of eigenvalues less than or equal to v is $O(v)$, so the number of j with $\varkappa_j \leqslant K$ is $O(K^2)$. By the calculation of Section 6 $\varkappa_j \geqslant 1$ for $j \geqslant 1$. Our notation here is slightly different from that of Kuznetsov and Iwaniec, as the function A absorbs a factor $2\sqrt{n}/\pi$. In the spectral theorem we are obliged to normalise a Maass wave form f (here we drop the suffix j) by $(f, f) = 1$. However

$$a(n) = \frac{b(n)\sqrt{n}}{b(1)}$$

is a multiplicative arithmetic function. The Ramanujan conjecture (extended to Maass wave forms) is that

$$|a(p)| \leqslant 2.$$

From the theory of Hecke operators it will be clear that

$$|a(p)| \leqslant p+1.$$

Bruggeman has discussed the size of the normalising factor $b^{(j)}(1)$, and has shown that it must vary somewhat with j to counteract the arithmetic properties of $a^{(j)}(n)$ for different f_j.

First we give a crude order of magnitude bound, which will be improved by the Kuznetsov theory.

GEOMETRIC LEMMA. *The line segment* $[-1/2+i\delta, 1/2+i\delta]$ *cuts* $\ll 1/\delta$ *copies of the fundamental domain* D *of* $\Gamma(1)$.

Proof. Suppose $z \in D$ and $\operatorname{Im} \gamma z \geqslant \delta$. Then

$$c^2 y + (cx+d)^2/y \leqslant 1/\delta,$$

so $c^2 \leqslant 2/\delta\sqrt{3}$. But now

$$2c\,|cx+d| \leqslant (cx+d)^2/y + c^2 y \leqslant 1/\delta,$$

and

$$|cx+d| \leqslant 1/2\delta c \ll 1/\sqrt{\delta}.$$

So c and d, which determine the coset representatives, each take $\ll 1/\sqrt{\delta}$ values. This nice argument was suggested by conversations with Dr Pleasants.

From the geometric lemma we see that

$$\int_{x=-1/2}^{1/2} \int_{y=\delta}^{2\delta} |f(x, y)|^2 \frac{dx\,dy}{y^2} \ll \frac{1}{\delta}(f, f) \ll \frac{1}{\delta},$$

when f is a normalised Maass wave form. The left-hand side is

$$\int_{\delta}^{2\delta} \sum_{-\infty}^{\infty} |b(m)|^2 A^2(2\pi m y) \frac{dy}{y^2}.$$

SADDLE POINT LEMMA. *With Δ large compared with $1/\varkappa$*

$$\int_0^{\infty} \cos(\varkappa x) \exp\left(-\frac{\operatorname{ch} x}{\Delta}\right) dx \sim \frac{\sqrt{(2\pi)}\cos(\varkappa \log 2\varkappa\Delta - \varkappa - \pi/4)}{\sqrt{(k\Delta)}\exp(\pi\varkappa/2)}.$$

Proof (sketched only). The saddle point nearest to the real axis is given by $\operatorname{sh} z_0 = i\varkappa\Delta$, $z_0 \sim i\pi/2 + \log 2\varkappa\Delta$.

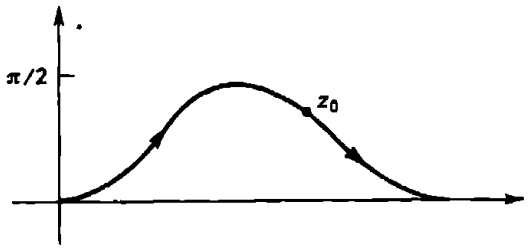

Fig. 30

Deform the contour to pass from 0 to $+\infty$ through z_0, with $z = z_0 + u\,\mathrm{cis}(-\pi/4)$; we want

$$\int \exp\left(i\varkappa z_0 - \frac{\operatorname{ch} z_0}{\Delta} + \frac{iu^2}{2\Delta}\operatorname{ch} z_0 - \ldots\right) du.$$

If u is real near $u = 0$, the main term is

$$\operatorname{Re}\left(\frac{\exp(i\varkappa z_0 - \operatorname{ch} z_0/\Delta)\sqrt{\pi}\ \operatorname{cis}(-\pi/4)}{|\tfrac{1}{2}\operatorname{ch} z_0|^{1/2}}\right).$$

COROLLARY. *If* $s = 1/2 + i\varkappa$, *and* $u/\varkappa$ *is small*

$$A(u) = \sqrt{\left(\frac{2u}{\pi}\right)}\int_0^\infty \exp(-u \operatorname{ch} x) \operatorname{ch} i\varkappa x\, dx \sim \frac{2u \cos(\varkappa \log 2\varkappa/ue - \pi/4)}{\sqrt{\varkappa}\exp(\pi\varkappa/2)}.$$

We now have

$$\frac{1}{\delta} \gg \sum_{m<1/\delta} |b(m)|^2 \int_\delta^{2\delta} A^2(2\pi m y)\frac{dy}{y^2} \gg \sum_{m<1/\delta} \frac{\delta m^2 |b(m)|^2}{\varkappa \exp \pi\varkappa}.$$

Two useful bounds can now be deduced:

$$|b(m)|^2 \ll \varkappa e^{\pi k} \quad \text{and} \quad \sum_1^M m|b(m)|^2 \ll \varkappa e^{\pi\varkappa} M \log M.$$

Further reading
BRUGGEMAN
JEFFREYS and JEFFREYS

11. Hecke operators

The ideas behind Hecke operators go back to Poincaré. Mordell used them to prove that Ramanujan's τ-function was multiplicative, and Hecke, Petersson, Atkin, Lehner, Rankin and others developed a systematic theory.

The first idea is the correspondence between automorphisms of the group and symmetries of its fundamental domain. All the congruence subgroups contain $\begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$ if they contain $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The fundamental domain can be so drawn as to be symmetric in the line $x = 0$, a left-right reflection. If $f(x, y)$ is automorphic, so is $g(x, y) = f(-x, y)$, and if f is a Maass wave form, so is g, and with the same eigenvalue. Now $\frac{1}{2}(f+g)$ and $\frac{1}{2}(f-g)$ span the same linear space as f and g, and they are orthogonal. Hence we can suppose that the functions of the orthonormal basis for cusp forms are either symmetric or skew-symmetric under this symmetry. In terms of the Fourier coefficients, symmetric means that

$$b(m) = b(-m) = \overline{b(m)},$$

so that $b(m)$ is real, f is a cosine series, and skew-symmetric means that

$$b(m) = -b(-m) = -\overline{b(m)},$$

so that $b(m)$ is purely imaginary, and f is a sine series. This symmetry is called the graph automorphism by students of Lie groups, and T_{-1} by Kuznetsov.

The congruence subgroups $\Gamma^0(N)$, $\Gamma_0(N)$ and $\Gamma(N)$ of the modular group posess further automorphisms. Recall that $\Gamma(N)$ consists of matrices congruent (mod N) to the identity, $\Gamma_0(N)$ and $\Gamma^0(N)$ consist of matrices respectively upper and lower triangular (mod N). Suppose that $N = QR$ with $(Q, R) = 1$. Let

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad W_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix},$$

with $A_i D_i - B_i C_i = Q$, $Q \mid A_i$, $Q \mid D_i$, $N \mid B_i$. For any γ of $\Gamma^0(N)$

$$\begin{aligned}(\text{adj } W_1)\gamma W_2 &= \begin{pmatrix} D_1 & -B_1 \\ -C_1 & A_1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \\ &\equiv \begin{pmatrix} 0 & 0 \\ -C_1 & 0 \end{pmatrix}\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 0 \\ C_2 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \pmod{Q}.\end{aligned}$$

Hence $W_1^{-1}\gamma W_2$ is an integer matrix, with top right entry divisible by R. In fact, since $\det W_1 = Q$ and $Q^2 \mid (A_1, B_1)(A_2, B_2)$, the top right entry is also divisible by Q, and so $W_1^{-1}\gamma W_2$ is in $\Gamma^0(N)$. We deduce that if $f(z)$ is automorphic for $\Gamma^0(N)$, so is $f \mid W$ defined by $(f \mid W)(z) = f(Wz)$, and $f \mid W$ depends only on the congruence conditions defining W, not on the particular choice W_i, and

$$f \mid W = f \mid \text{adj } W, \quad f \mid W^2 = f \mid W \,\text{adj } W = f.$$

Hence W gives a Hermitian operator $W_Q : f \to f \mid W$ with eigenvalues ± 1. We can exhibit an explicit W as

$$W = (\text{adj } A_Q) V_{R,Q},$$

where

$$A_Q = \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}, \quad V_{R,Q} = \begin{pmatrix} \bar{Q} & R \\ -\bar{R} & Q \end{pmatrix}$$

where $R\bar{R} + Q\bar{Q} = 1$. When $-Q$ is a quadratic residue mod R, there will be a matrix W which represents a rotation by π. For instance $\begin{pmatrix} 0 & N \\ -1 & 0 \end{pmatrix}$ is a W_N representing a rotation about $i\sqrt{N}$, used in Section 5 with $N = 13$.

The second principle is that if R is any rigid motion, and $f(z)$ is automorphic under a group Γ, then $f(R^{-1}z)$ is automorphic under the group $R\Gamma R^{-1}$. This group is said to be commensurable with Γ if the subgroup $\Gamma \cap R\Gamma R^{-1}$ is actually of finite index in Γ (and thus also in $R\Gamma R^{-1}$). In particular

$$A_N \Gamma^0(N) A_N^{-1} = \Gamma_0(N),$$

$$A_N \Gamma^0(N^2) A_N^{-1} = \Gamma_0(N) \cap \Gamma^0(N) \geqslant \Gamma(N).$$

The fundamental domains and the eigenfunctions of $\Gamma^0(N)$ and $\Gamma_0(N)$ differ only by a trivial coordinate change $z \to z/N$. This is the basic concept in Atkin and Lehner's beautiful theory of oldforms and newforms.

The third principle is that if H is a subgroup of G of finite index, so that G is a finite coset union $\bigcup H\gamma_i$, and if f is automorphic on H, then $\sum f(\gamma_i z)$ is automorphic on G. We apply this principle with $G = \Gamma^0(M)$, $H = \Gamma^0(MN)$. Suppose $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma^0(M)$. Let $Q = (a, N)$, $N = QR$, $a = Qe$. Then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u & -BM \\ t & Q \end{pmatrix} \in \Gamma^0(MN)$$

when

$$Be \equiv b/M \pmod{R}.$$

Since $(e, R) = 1$, B lies in an unique residue class (mod R), and since $(a, b) = 1$ there is an integer in this residue class with $(Q, BM) = 1$: we choose B to be the least non-negative such integer. We also choose t and u so that

$$Qu + BMt = 1.$$

Then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma^0(MN) \begin{pmatrix} Q & BM \\ -t & u \end{pmatrix}.$$

So each $Q \mid N$ gives at most N/Q coset representatives (if $(Q, R) > 1$, then B has to lie in a residue class mod R that is prime to (Q, R)).

If $f(z)$ is automorphic for $\Gamma^0(M)$, $f(A_N z)$ is automorphic for $\Gamma^0(MN)$, and

$$f\left(A_N \begin{pmatrix} Q & BM \\ -t & u \end{pmatrix} z\right) = f\left(\begin{pmatrix} MRt+1 & M \\ Rt & 1 \end{pmatrix} A_N \begin{pmatrix} Q & BM \\ -t & u \end{pmatrix} z\right)$$

$$\sim = f\left(\begin{pmatrix} Q & MR+BM \\ 0 & R \end{pmatrix} z\right) = f\left(\begin{pmatrix} Q & BM \\ 0 & R \end{pmatrix} z\right).$$

Hence

$$h(z) = \sum_{\substack{QR=N \\ (Q,M)=1}} \sum_{\substack{B \bmod R \\ (B,Q,R)=1}} f\left(\begin{pmatrix} Q & BM \\ 0 & R \end{pmatrix} z\right)$$

is also on $\Gamma^0(M)$. The condition of summation $(B, Q, R) = 1$ is inconvenient here. If $(B, Q, R) = d$, then $B = bd$, $Q = dq$, $R = dr$, $N = d^2n$, and

$$\sum_{\substack{QR=N \\ (Q,M)=1}} \sum_{\substack{B \bmod R \\ (B,Q,R)=d}} f\left(\begin{pmatrix} Q & BM \\ 0 & R \end{pmatrix} z\right) = \sum_{\substack{qr=n \\ (q,M)=1}} \sum_{\substack{b \bmod r \\ (b,q,r)=1}} f\left(\begin{pmatrix} q & bM \\ 0 & r \end{pmatrix} z\right).$$

This is a sum of the same type with N replaced by $n = N/d^2$. Summing over all d with $d^2 \mid N$, $(d, M) = 1$ we see that

$$g(z) = \sum_{\substack{QR=N \\ (Q,M)=1}} \sum_{B \bmod R} f\left(\begin{pmatrix} Q & BM \\ 0 & R \end{pmatrix} z\right)$$

is also on $\Gamma^0(M)$. If

$$f(z) = f_n(y)\, e(nx/M),$$

then

$$g(z) = \sum_{\substack{QR=N \\ (Q,M)=1}} \sum_n f_n\left(\frac{Qy}{R}\right) \sum_{B \bmod R} e\left(\frac{Qnx + BMn}{RM}\right)$$

$$= \sum_{\substack{QR=N \\ (Q,M)=1}} R \sum_m f_{mR}\left(\frac{Qy}{R}\right) e\left(\frac{mQx}{M}\right).$$

In the special case when $f(z)$ is a cusp form,

$$f_n(y) = b(n)\, A(2\pi ny/M),$$

and

$$g(z) = \sum_{\substack{QR=N \\ (Q,M)=1}} R \sum_m b(mR)\, e\left(\frac{mQx}{M}\right) A\left(\frac{2\pi mQy}{M}\right).$$

The map from $f(z)$ to $g(z)$ is a Hecke operator on $\Gamma^0(M)$. In Atkin and Lehner's notation

$$g(z) = \begin{cases} T_N f(z) & \text{if } (M, N) = 1, \\ U_N f(z) & \text{if all prime factors of } N \text{ also divide } M. \end{cases}$$

Hecke operators derived from different subgroups $\Gamma^0(MN)$ of $\Gamma^0(M)$ commute, and in fact

$$T_m T_n = \sum_{\substack{d \mid m \\ d \mid n}} d\, T_{mn/d^2},$$

which can be checked either on calculations with 2×2 matrices, or on the Fourier expansions of automorphic functions. In the case $(M, N) = 1$ a cunning choice of coset representatives shows that T_N is Hermitian. In fact the T_N with $(M, N) = 1$ and the W_Q with $QR = M$, $(Q, R) = 1$ form a commuting system of Hermitian operators on the automorphic forms, which map the finite dimensional space of cusp forms on $\Gamma^0(M)$ with a fixed eigenvalue into itself. We may choose a basis for which all these operators are diagonal, so that

$$W_Q f = \pm f, \quad T_N f = \frac{Nb(N)}{b(1)} f = a(N) N^{1/2} f,$$

say, where $a(N)$ is a multiplicative function. The Ramanujan Petersson conjecture is that $|a(p)| \leqslant 2$ when p is prime; Deligne has proved the original form of the conjecture for holomorphic cusp forms.

Atkin and Lehner in their classic paper prove a result which translates into our situation as follows. Let V be a space of simultaneous eigenfunctions of the Laplacian on $\Gamma^0(M)$ and of all T_N with $(M, N) = 1$. If V has dimension greater than one, it is spanned by 'oldforms' $f(z/Q)$, where $Q > 1$, $Q \mid M$ and $f(z)$ is automorphic on $\Gamma^0(M/Q)$, a larger group. If the dimension is one, the generator $f(z)$ of V is a 'newform', not induced from any larger subgroup of this type. It has $b(1) \neq 0$. Their only result which does not go through immediately in our context is the Determination Theorem (Bestimmungssatz) that every oldform is derived from an unique newform on a larger group. W. W. Li gave another proof which does work for Maass wave forms.

Further reading

ATKIN and LEHNER
RANKIN

12. Kuznetsov's Basic Lemma

So far we have only written down explicit automorphic functions in the "Eisenstein space" of functions orthogonal to the cusp forms. The more general construction, due originally to Poincaré, takes

$$U_n(z) = \sum_{\langle T\rangle\gamma} e(n \operatorname{Re} \gamma z) g(\operatorname{Im} \gamma z).$$

By suitable choice of the function g (depending on n) Bruggeman makes $U_n(z)$ an eigenfunction of the Laplacian with prescribed eigenvalue $\lambda = s(1-s)$. For general s there will be no square-integrable eigenfunction with this eigenvalue; and even if there is one, it is not equal to Bruggeman's function. Kuznetsov takes $n > 0$ and

$$g(y) = y^s e^{-2\pi n y}, \quad U_n(z, s) = \sum_{\langle T\rangle\gamma} e(n\gamma z)(\operatorname{Im} \gamma z)^s.$$

By comparison with the Eisenstein series, Kuznetsov's Poincaré series converges absolutely for $\operatorname{Re} s > 1$. It satisfies

$$\operatorname{div} \operatorname{grad} U_n(z, s) = -s(1-s)\,U_n(z, s) - 4\pi n s U_n(z, s+1).$$

The first step is to work out the Fourier series

$$U_n(z, s) = \sum_m a_m(y)\, e(mx).$$

In Bruggeman's case the coefficients $a_m(y)$ come out in terms of Bessel functions; this calculation was done by Niebur. In our case

$$a_m = \delta_{mn} y^s e^{-2\pi n y} + \int_{-1/2}^{1/2} e(-mx) \sum_{c=1}^{\infty} \sum_{\substack{d=-\infty \\ (c,d)=1}}^{\infty} \frac{y^s}{|cz+d|^{2s}}\, e\left(\frac{an}{c} - \frac{n}{c(cz+d)}\right) dx.$$

As outlined in Section 3, we simplify the integral by collecting terms into residue classes for d modulo c, then making a variable change $u = x + d/c$. The integral becomes

$$\sum_{c=1}^{\infty} \sum_{d \bmod c}^{*} \int_{-\infty}^{\infty} e\left(\frac{an}{c}\right) e\left(-nx - \frac{n}{c(cz+d)}\right) \frac{y^s}{|cz+d|^{2s}}\, dx$$

$$= \sum_{c=1}^{\infty} \sum_{d \bmod c}^{*} e\left(\frac{an}{c} + \frac{dm}{c}\right) \int_{-\infty}^{\infty} e\left(-mu - \frac{n}{c^2(u+iy)}\right) \frac{y^s}{c^{2s}(u^2+y^2)^s}\, du.$$

Another advantage of working over the full modular group now appears. The sum over d is the Kloosterman sum

$$S(n, m; c) = \sum_{\substack{a \bmod c \\ ad \equiv 1 (\bmod c)}} \sum_{d \bmod c} e\left(\frac{an}{c} + \frac{dm}{c}\right) = S(m, n; c),$$

which, as André Weil has proved, satisfies

$$S(m, n; c) \ll c^{1/2} \tau(c) \ll c^{1/2+\varepsilon}.$$

Since

$$\left| e\left(-\frac{n}{c^2(u+iy)}\right) \right| \leqslant 1,$$

the individual terms of the Fourier series converge like

$$\sum_1^{\infty} \frac{|S(m, n; c)|}{c^{2\sigma}},$$

where as usual $s = \sigma + it$. Moreover if we treat the integral over u as a contour integral, and take it to

$$\text{Im } u = -y/2 \quad \text{for } m \geqslant 0, \qquad \text{Im } u = y/2 \quad \text{for } m < 0$$

then

$$|e(-mu)| < e^{-\pi|m|y}$$

and the other estimates still hold, so that the Fourier series itself converges. Hence for $\text{Re } s > 3/4$ the Fourier series gives an analytic continuation of $U_n(z, s)$.

Poincaré series behave like Eisenstein series in inner products:

$$\begin{aligned}(U_n(z, s), f(z)) &= \int\int_D U_n(z, s)\overline{f(z)}\, d\mu \\ &= \int_0^\infty \int_{-1/2}^{1/2} y^s e(nx) e^{-2\pi ny} \overline{f(z)}\, \frac{dx\, dy}{y^2} \\ &= \int_0^\infty y^s e^{-2\pi ny} \overline{f_n(y)}\, \frac{dy}{y^2}.\end{aligned}$$

If $f(z)$ happens to be an eigenfunction

$$f_n(y) = b_n A(2\pi ny) = \pi b_n \sqrt{(ny)}\, \text{Kh}_{i\varkappa}(2\pi ny).$$

The integral is one worked out in the section on Bessel functions, with $2\pi ny = u$. Hence

$$(U_n(z, s), f(z)) = \frac{\overline{b(n)}\, \underline{|s-\frac{3}{2}+i\varkappa}\ \underline{|s-\frac{3}{2}-i\varkappa}}{(4\pi n)^{s-1}\, \underline{|s-1}}.$$

The integral also converges when $f(z)$ is an Eisenstein series, for which we know $b(n)$ explicitly

$$(U_n(z, s), E(z, \tfrac{1}{2}+i\varkappa)) = \frac{\pi^{3/2-i\varkappa-s}\tau_{i\varkappa}(n)\, \underline{|s-\frac{3}{2}+i\varkappa}\ \underline{|s-\frac{3}{2}-i\varkappa}}{4^{s-1} n^{s-1/2} \zeta(1-2i\varkappa)\, \underline{|-\frac{1}{2}-i\varkappa}\ \underline{|s-1}}.$$

Although $U_n(z, s)$ is not of compact support, the inner product with $E(z, 1/2+i\varkappa)$ exists, since the Eisenstein series is bounded by a power of y as $y \to \infty$, whilst the summand $y^s e(nz)$ in the definition of U_n is exponentially small for large y.

Kuznetsov's strategy is to express an inner product $(U_m(z, s), U_n(z, \bar{w}))$ of two Poincaré series in terms of its components in the spectral

decomposition

$$(U_m(z, s), f)(f, U_n(z, \bar{w})) = \frac{\overline{b(m)}\, b(n)}{(4\pi m)^{s-1}(4\pi n)^{w-1}} \Lambda(s, w; \varkappa),$$

where $\Lambda(s, w; \varkappa)$ is the product

$$\Lambda(s, w; \varkappa) = \frac{\lfloor s-\frac{3}{2}+i\varkappa \;\; \lfloor s-\frac{3}{2}-i\varkappa \;\; \lfloor w-\frac{3}{2}+i\varkappa \;\; \lfloor w-\frac{3}{2}-i\varkappa}{\lfloor s-1 \;\; \lfloor w-1},$$

so that as $\varkappa \to \infty$

$$|\Lambda(s, w; \varkappa)| \sim \frac{(2\pi)^2 \varkappa^{2\mathrm{Re}\, s + 2\mathrm{Re}\, w - 4} e^{-2\pi\varkappa}}{\lfloor s-1 \;\; \lfloor w-1}.$$

By the coefficient bound for Maass wave forms

$$|\overline{b(m)}\, b(n)| \ll \varkappa e^{\pi\varkappa}.$$

From the proof of the spectral theorem $\varkappa_j^2 \gg j$, so the sum over eigenfunctions converges exponentially. When f is replaced by the Eisenstein series $E(z, 1/2+it)$, the product of coefficients $\overline{b(m)}\, b(n)$ is replaced by

$$\frac{\pi^{1/2-it}\sqrt{m}\,\tau_{it}(m)\,\pi^{1/2+it}\sqrt{n}\,\tau_{it}(n)}{\lfloor -1/2-it \;\; \zeta(1-2it) \lfloor -1/2+it \;\; \zeta(1+2it)},$$

and from the functional equation of the factorial function

$$\frac{1}{\lfloor -1/2-it \;\; \lfloor -1/2+it} = \frac{\mathrm{ch}\, \pi t}{\pi}.$$

We can now write

$$(U_m(z, s), U_n(s, \bar{w})) = \sum_{j=1}^{\infty} \frac{\overline{b_j(m)}\, b_j(n) \Lambda(s, w; \varkappa_j)}{(4\pi m)^{s-1}(4\pi m)^{w-1}} +$$

$$+\frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{\tau_{it}(m)\tau_{it}(n)\Lambda(s, w; t)\,\mathrm{ch}\, \pi t}{\sqrt{(mn)}\,\zeta(1+2it)\zeta(1-2it)(4\pi m)^{s-1}(4\pi n)^{w-1}}\, dt.$$

There is no term $j = 0$, since the constant function has no term in $e(mx)$ in its Fourier series.

The sum over j in the expression for the inner product is convergent for any complex s and w, being meromorphic in s for fixed w, and meromorphic in w for fixed s. The integral is subtler, as poles can lie on the contour. It is regular in s and w for $\mathrm{Re}\, s > 1/2$, $\mathrm{Re}\, w > 1/2$, and can be continued to the left by the calculus of residues. If $s+w = 2$ there is a great simplification in

$\Lambda(s, w; t)$. Putting $s = 1+iu$, $w = 1-iu$

$$\Lambda(s, w; t) = \frac{\underline{|iu-1/2+it}\ \underline{|iu-1/2-it}\ \underline{|-iu+1/2+it}\ \underline{|-iu-1/2-it}}{\underline{|iu}\ \underline{|-iu}}$$

$$= \frac{\pi}{\operatorname{ch} \pi(t+u)} \frac{\pi}{\operatorname{ch} \pi(t-u)} \frac{\sin \pi iu}{\pi iu} = \frac{\pi}{u} \frac{\operatorname{sh} \pi u}{\operatorname{ch} \pi t} K(t, u),$$

where $K(t, u)$ (called $H(t, u)$ by Kuznetsov) is the kernel

$$K(t, u) = \frac{\operatorname{ch} \pi t}{\operatorname{ch} \pi(t+u) \operatorname{ch} \pi(t-u)}.$$

Since $U_m(z, s)$ is a Poincaré series,

$$(U_m(z, s), U_n(z, \bar{w})) = \int_{y=0}^{\infty} y^s e^{-2\pi my} \bar{a}_m(y) \frac{dy}{y^2},$$

where $a_m(y)$ is the Fourier coefficient of $U_n(z, \bar{w})$:

$$\delta_{mn} y^{\bar{w}} e^{-2\pi my} + \sum_{c=1}^{\infty} \frac{S(m, n; c)}{c^{2\bar{w}}} \int_{u=-\infty}^{\infty} e\left(-mu - \frac{n}{c^2(u+iy)}\right) \frac{y^{\bar{w}}}{(u^2+y^2)^{\bar{w}}} du.$$

The first term integrates to

$$\delta_{mn}(4\pi m)^{1-s-w} \underline{|s+w-1}.$$

In the second term the inner integral over u is $\ll y^{1-\operatorname{Re} w} e^{\pi - \pi my}$, uniformly in c, so that we are justified in summing over c last for $\operatorname{Re}(s-w+1) > 1$. When we put $u = -ty$

$$\int_{y=0}^{\infty} y^s e^{-2\pi my} \int_{u=-\infty}^{\infty} \exp\left(2\pi imu + \frac{2\pi in}{c^2(u-iy)}\right) \frac{y^w}{(u^2+y^2)^w} \frac{du\, dy}{y^2}$$

$$= \int_{y=0}^{\infty} \frac{y^{s-w-1}}{e^{2\pi my}} \int_{t=-\infty}^{\infty} \exp\left(-2\pi imty - \frac{2\pi n}{c^2 y(1-it)}\right) \frac{1}{(t^2+1)^w} dt\, dy.$$

For $1 < \operatorname{Re} w < \operatorname{Re} s$ the double integral converges absolutely, so we may integrate over y first. Now

$$\mathrm{Kh}_{w-s}(x) = \frac{1}{\pi} \int_0^{\infty} \exp\left(-\frac{x}{2}\left(u + \frac{1}{u}\right)\right) \frac{du}{u^{w-s+1}}.$$

If we put

$$x = 4\pi\sqrt{\left(\frac{mn}{c^2}\frac{1+it}{1-it}\right)}, \qquad u = \sqrt{\left(\frac{mn}{c^2}(1+t^2)y^2\right)},$$

we can identify the integral over y as a Bessel function, so that the double integral takes the form

$$\int_{t=-\infty}^{\infty} \frac{\pi}{(t^2+1)^w}\left(\frac{n}{mc^2(t^2+1)}\right)^{(s-w)/2} \mathrm{Kh}_{w-s}\left(4\pi\sqrt{\left(\frac{mn}{c^2}\frac{1+it}{1-it}\right)}\right) dt.$$

Kuznetsov's ingenuity has, however, only just begun. Writing

$$\alpha(c) = 4\pi\sqrt{\left(\frac{mn}{c^2}\right)}, \qquad v = \sqrt{\left(\frac{1-it}{1+it}\right)},$$

so that

$$v+\frac{1}{v} = \frac{2}{(1+t^2)}, \qquad \frac{2dv}{v} = \frac{-2idt}{1+t^2},$$

we see that

$$\left(U_m(z, s), U_n(z, \bar{w})\right) = \delta_{mn}(4\pi m)^{1-s-w}\underline{\lfloor s+w-1} +$$

$$+\sum_{c=1}^{\infty}\frac{S(m, n; c)}{c^{s+w}}\left(\frac{n}{m}\right)^{(s-w)/2}\pi i\int_C \{\tfrac{1}{2}(v+1/v)\}^{s+w-2}\mathrm{Kh}_{w-s}(4\pi v\alpha(c))\frac{dv}{v},$$

where the contour C is the right-hand half of the unit circle, from $-i$ to i, taken anticlockwise. Now

$$\mathrm{Kh}_{w-s}(z) \ll |z|^{\mathrm{Re}(w-s)}\log(1/|z|) \quad \text{as } z \to 0.$$

We are still assuming Re $w <$ Re s here, but there is now symmetry in w and s, so we can continue analytically to Re $s > 3/4$, Re $w > 3/4$ (this requires Weil's bound for the Kloosterman sum to give convergence in c). Again we shall take $s+w = 2$, $s = 1+iu$. To get convergence in the sum over c, any non-trivial exponent in the bound for the Kloosterman sum will do, letting u lie on a neighbourhood of the real axis.

We now have two expressions for a meromorphic function $F(s, w)$ of two complex variables, with

$$F(s, w) = \left(U_m(z, s), U_n(z, \bar{w})\right)$$

for $\operatorname{Re} s > 1$, $\operatorname{Re} w > 1$. We have just shown that

$$F(1+iu, 1-iu) = \frac{\delta_{mn}}{4\pi m} + \sum_{c=1}^{\infty} \frac{S(m, n; c)}{c^2} \left(\frac{n}{m}\right)^{iu} \int_C \mathrm{Kh}_{2iu}(v\alpha(c)) \frac{dv}{v}.$$

We now deform the contour C.

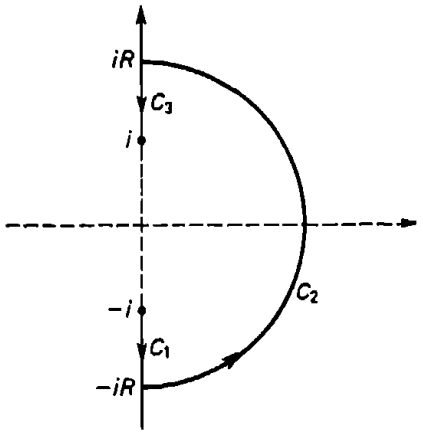

Fig. 31

The deformed contour consists of
C_1: line segment $-i$ to $-iR$,
C_2: semicircle centre 0 radius R in $\operatorname{Re} v > 0$,
C_3: line segment iR to i.
On C_2 we have

$$\mathrm{Kh}_{2iu}(v\alpha(c)) \ll |v|^{-1/2},$$

so that the integral over C_2 tends to zero. On C_1 we put $v = -iy$, so that

$$\begin{aligned}\sin 2\pi iu\ \mathrm{Kh}_{2iu}(-i\alpha y) &= I_{2iu}(-i\alpha y) - I_{-2iu}(-i\alpha y) \\ &= \operatorname{cis}(-\pi iu) J_{2iu}(\alpha y) - \operatorname{cis}(\pi iu) J_{-2iu}(\alpha y).\end{aligned}$$

Similarly we put $v = iy$ on C_3 and

$$\sin(2\pi iu)\ \mathrm{Kh}_{2iu}(i\alpha y) = \operatorname{cis}(\pi iu) J_{2iu}(\alpha y) - \operatorname{cis}(-\pi iu) J_{-2iu}(\alpha y).$$

Subtracting, we obtain for the integrals along C_1 and C_3

$$\frac{1}{\sin 2\pi iu} \int_{y=1}^{\infty} (-2i \sin \pi iu)\left(J_{2iu}(\alpha y) + J_{-2iu}(\alpha y)\right) \frac{dy}{y}$$

$$= \frac{-i}{\operatorname{ch} \pi u} \int_{t=\alpha}^{\infty} \left(J_{2iu}(t) + J_{-2iu}(t)\right) \frac{dt}{t},$$

so that

$$F(1+iu,\ 1-iu) = \frac{\delta_{mn}}{4\pi n} + \sum_{c=1}^{\infty} \frac{S(m,\ n;\ c)}{c^2} \left(\frac{n}{m}\right)^{iu} \frac{\pi}{\operatorname{ch} \pi u} \int_{\alpha(c)}^{\infty} \left(J_{2iu}(t) + J_{-2iu}(t)\right) \frac{dt}{t}.$$

We now obtain Kuznetsov's first result by equating our two expressions for the analytic continuation of the product of two Poincaré series.

BASIC LEMMA. *For* $-1/4 < \operatorname{Im} u < 1/4$

$$\frac{\delta_{mn}}{\pi} \frac{u}{\operatorname{sh} \pi u} + \sum_{c=1}^{\infty} \frac{S(m,\ n;\ c)}{c} \frac{2u}{\operatorname{sh} 2\pi u} \alpha(c) \int_{\alpha(c)}^{\infty} \left(J_{2iu}(t) + J_{-2iu}(t)\right) \frac{dt}{t}$$

$$= \sum_{j=1}^{\infty} \frac{4\pi \sqrt{mn}\ \overline{b_j(m)}\, b_j(n)}{\operatorname{ch} \pi \varkappa_j} K(\varkappa_j,\ u) + \int_{-\infty}^{\infty} \frac{\tau_{it}(m)\, \tau_{it}(n)\, K(t,\ u)}{\zeta(1+2it)\, \zeta(1-2it)}\, dt.$$

To complete the proof of the Basic Lemma we observe that both sides are equal to

$$4\sqrt{mn} \frac{u}{\operatorname{sh} \pi u} \left(\frac{m}{n}\right)^{iu} F(1+iu,\ 1-iu).$$

Further reading
BRUGGEMAN
KUZNETSOV

13. The First Trace Formula

The Kuznetsov trace formulae connect Kloosterman sums ('known') with eigenvalues $1/4+\varkappa_j^2$ ('unknown'). The first formula has a specified function $h(\varkappa_j)$ on the unknown side, and integral transforms of $h(z)$ on the known side. The second formula has a specified weight function $g(x)$ multiplying the Kloosterman sums, and integral transforms of $g(x)$ on the unknown side. The Second Trace Formula is the one used to estimate number-theoretical sums.

The function $h(z)$ is brought in by means of the following lemma.

STIELTJES' LEMMA. *If* $h(z)$ *is regular in the strip* $-1/2 \leqslant y \leqslant 1/2$ *and* $\int_{-\infty}^{\infty} e^{-\pi|x|} |h(x+iy)|\, dx$ *converges for* $-1/2 \leqslant y \leqslant 1/2$, *then*

$$\int_{u=-\infty}^{\infty} \tfrac{1}{2} \left\{ h\left(u+\frac{i}{2}\right) + h\left(u-\frac{i}{2}\right) \right\} \frac{du}{\operatorname{ch} \pi u} = h(0).$$

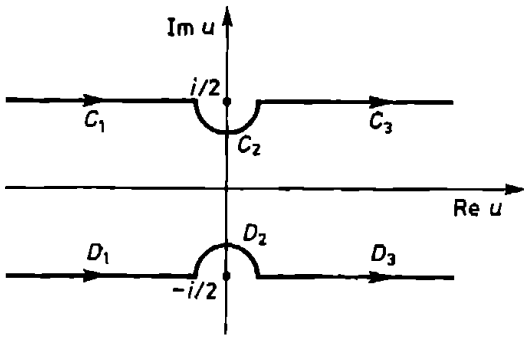

Fig. 32

Proof. Move the line of integration for the term $h(u-i/2)$ to C made up of

C_1: line segment $-\infty+i/2$ to $-\delta+i/2$,

C_2: anticlockwise semicircle centre $i/2$ radius δ from $-\delta+i/2$ to $\delta+i/2$

C_3: line segment $\delta+i/2$ to $\infty+i/2$

and the line of integration for the term $h(u+i/2)$ to D:

D_1: line segment $-\infty-1/2i$ to $-\delta-1/2i$,

D_2: clockwise semicircle centre $-1/2i$ from $-\delta-1/2i$ to $\delta-1/2i$,

D_3: line segment $\delta-1/2i$ to $\infty-1/2i$.

Since $\operatorname{ch} \pi(u+i/2) = -\operatorname{ch} \pi(u-i/2)$, the integral along C_1 cancels that along D_1, the integral along C_3 cancels that on D_3, and the integrals on C_2 and D_2 add to a circle round the pole of $\operatorname{ch} \pi u$ at $u = i/2$, with residue $1/\pi i$.

COROLLARY 1. *For real* t

$$\int_{-\infty}^{\infty} \frac{\operatorname{ch} \pi t \operatorname{ch} \pi u}{\operatorname{ch} \pi(t+u) \operatorname{ch} \pi(t-u)} \{h(u+i/2)+h(u-i/2)\}\, du = h(t)+h(-t).$$

Proof. We use the partial fraction decomposition of Kuznetsov's kernel $K(t, u)$:

$$\frac{1}{\operatorname{ch} \pi(t+u)}+\frac{1}{\operatorname{ch} \pi(t-u)} = \frac{2 \operatorname{ch} \pi t \operatorname{ch} \pi u}{\operatorname{ch} \pi(t+u) \operatorname{ch} \pi(t-u)}.$$

COROLLARY 2. *For* $\operatorname{Re} s > 1/2$

$$\frac{1}{\pi} \int_{u=-\infty}^{\infty} u \operatorname{th} \pi u \, \underline{\lfloor s-3/2+iu} \, \underline{\lfloor s-3/2-iu} \sin \pi s \, du = \frac{\underline{\lfloor s-1}}{\underline{\lfloor -s}}.$$

Proof. We put

$$h(u) = \frac{\underline{\lfloor s-1+iu}}{\underline{\lfloor -s+iu}} = \underline{\lfloor s-1+iu} \, \underline{\lfloor s-1-iu} \sin(\pi s - i\pi u).$$

After some calculation we find

$$h(u+i/2)+h(u-i/2)$$
$$= -\frac{2iu}{\pi}\,\underline{\lfloor s-3/2+iu}\;\underline{\lfloor s-3/2-iu}\;\{\cos \pi s \operatorname{ch} \pi u + i \sin \pi s \operatorname{sh} \pi u\}.$$

The integral with $u \operatorname{ch} u$ is 0 by symmetry, so we have the result.

A similar calculation without a residue evaluation gives

$$\int_{-\infty}^{\infty} \frac{u \operatorname{ch} \pi u}{2 \operatorname{sh} \pi u}\{h(u+i/2)+h(u-i/2)\}\,du = \int_{-\infty}^{\infty} \frac{v \operatorname{sh} \pi v}{\operatorname{ch} \pi v} h(v)\,dv.$$

We are now ready to multiply the Basic Lemma by

$$\tfrac{1}{2} \operatorname{ch} \pi u\{h(u+i/2)+h(u-i/2)\}$$

and integrate over u. Since

$$J_{2iu}(t) \ll \frac{e^{\pi u}}{(u^2+t^2)^{1/4}},$$

the double integral

$$\int_{u=-\infty}^{\infty} \tfrac{1}{2} \operatorname{ch} \pi u \frac{2u}{\operatorname{sh} \pi u} \int_{x}^{\infty} \{J_{2iu}(t)+J_{-2iu}(t)\}\frac{dt}{t}\{h(u+i/2)+h(u-i/2)\}\,du$$

converges absolutely like

$$\int_{-\infty}^{\infty} \int_{x}^{\infty} \frac{u|h(u+iy)|\,du\,dt}{(u^2+t^2)^{1/4}t}$$

and we may interchange orders of integration. By symmetry

$$\int_{-\infty}^{\infty} \frac{u}{2 \operatorname{sh} \pi u}\{h(u+i/2)+h(u-i/2)\}\{J_{2iu}(t)+J_{-2iu}(t)\}\,du$$
$$= 2\int_{-\infty}^{\infty} \frac{u}{2 \operatorname{sh} \pi u}\{h(u+i/2)+h(u-i/2)\}J_{2iu}(t)\,du.$$

We now move each term to the real axis: if $u = v-i/2 = -\frac{1}{2}i(2iv+1)$, then $\operatorname{sh} \pi u = -i \operatorname{ch} \pi v$, and if $u = v+i/2 = -\frac{1}{2}i(2iv-1)$, then $\operatorname{sh} \pi u = i \operatorname{ch} \pi v$.

The integral over u is

$$\int_{-\infty}^{\infty} \frac{-\frac{1}{2}i(2iv+1)h(v)}{-i \operatorname{ch} \pi v} J_{2iv+1}\, dv + \int_{-\infty}^{\infty} \frac{-\frac{1}{2}i(2iv-1)h(v)}{-i \operatorname{ch} \pi v} J_{2iv-1}\, dv$$

$$= -\int_{-\infty}^{\infty} \frac{h(v)}{\operatorname{ch} \pi v} 2ivt \frac{d}{dt} \frac{J_{2iv}(t)}{t}\, dv$$

by the recurrence relation

$$(s+1)J_{s+1}(t) - (s-1)J_{s-1}(t) = -2st \frac{d}{dt} \frac{1}{t} J_s(t),$$

and we can integrate over t in the double integral to get

$$2i \int_{-\infty}^{\infty} \frac{vh(v)}{\operatorname{ch} \pi v} J_{2iv}(\alpha)\, dv.$$

We wish to multiply by $\alpha S(m, n; c)/c$ with $\alpha = \alpha(c) = 4\pi\sqrt{(mn/c^2)}$ and sum. As $\alpha \to 0+$ we have for any σ in $0 < \sigma < 1/2$

$$\int_{-\infty}^{\infty} \frac{vh(v)}{\operatorname{ch} \pi v} J_{2iv}(\alpha)\, dv = \int_{t=-\infty}^{\infty} \frac{(t-i\sigma)h(t-i\sigma)}{\operatorname{ch} \pi(t-i\sigma)} J_{2it+2\sigma}(\alpha)\, dt \ll \alpha^{2\sigma}/(1-2\sigma),$$

where we have used the continuation of $h(v)$ off the real axis. The series over c converges like $\sum |S(m, n; c)|/c^{1+2\sigma}$, and any non-trivial exponent in the upper bound for the Kloosterman sum justifies the absolute convergence. For simplicity we suppose that $h(z) = h(-z)$. The other terms in the Basic Lemma are handled by Stieltjes' Lemma and its by-blows. We deduce Kuznetsov's First Trace Formula (or Sum Formula).

THEOREM. *Let $h(z)$ be regular for $|y| \leqslant 1/2$ and tend to zero sufficiently fast for $|x| \to +\infty$. Then*

$$\frac{\delta_{mn}}{\pi^2} \int_{-\infty}^{\infty} v \operatorname{th} \pi v\, h(v)\, dv + \frac{2i}{\pi} \sum_{c=1}^{\infty} \frac{S(m, n; c)}{c} \int_{-\infty}^{\infty} \frac{vh(v)}{\operatorname{ch} \pi v} J_{2iv}\left(4\pi\sqrt{(mn/c^2)}\right) dv$$

$$= \sum_{j=1}^{\infty} \frac{4\sqrt{mn}\, \overline{b_j(m)}\, b_j(n)\, h(\kappa_j)}{\operatorname{ch} \pi\kappa_j} + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\tau_{it}(m)\tau_{it}(n)h(t)}{\zeta(1+2it)\zeta(1-2it)}\, dt.$$

The infamous Selberg Trace Formula also contains the first integral on the left. The second term on the left is quite different in the Selberg formula, as the group elements are combined into conjugacy classes by Selberg, but into double cosets of the transvection group fixing $i\infty$ by Kuznetsov. On the right-hand side the Selberg formula has $\sum h(\varkappa_j)$ and a corresponding integral. The relation between the Kuznetsov and Selberg formulae still has to be cleared up. Kuznetsov has announced in Doklady ANSSR that he can deduce the Selberg formula and improved versions of its consequences. Zagier has a proof of the Selberg formula in the harder case when the fundamental region has cusps, which follows Kuznetsov in considering a Fourier coefficient. Finally there is Selberg's Jugendtraum: to deduce the Voronoi Summation Formula from the Selberg Trace Formula or something like it.

For the convergence in v we certainly need $\int_0^\infty x|h(x+iy)|\,dx$ to converge. The sum over j imposes a stronger condition. Note that if there were an eigenvalue $\lambda < 1/4$, instead of $\lambda = 1/4+\varkappa^2$ we would write $\lambda = \beta(1-\beta)$, and instead of $h(\varkappa)$ we would have $h(i(\beta-1/2))$. In the Selberg formula there is a term in $j=0$, $\lambda_0 = 0$, $\beta_0 = 1$, which explains why $h(z)$ has to be defined for $|y| \leqslant 1/2$. We have shown so far that $\varkappa_j \gg \sqrt{j}$ and that $b_j(m) \ll \sqrt{\varkappa_j}\exp \frac{1}{2}\pi\varkappa_j$. Hence we want $\sum \varkappa_j h(\varkappa_j)$ to converge, which requires $\int_0^\infty x^2 \max_{X>x}|h(X+iy)|\,dx$ to converge. However Kuznetsov has thoughtfully proved a mean square theorem

$$\sum_{\varkappa_j \leqslant \varkappa} \frac{4n|b_j(n)|^2}{\operatorname{ch} \pi\varkappa_j} = \frac{K^2}{\pi^2} + O\big(K(d^2(n)+\log K)\big) + O\big(\sqrt{nd_3(n)}\,\log^2 n\big),$$

where $d(n)$ and $d_3(n)$ are the ordinary divisor functions. This lets us relax the condition on $h(z)$ to $\int_0^\infty x \max_{X>x}|h(X+iy)|\,dx$ convergent.

The idea is to multiply the Basic Lemma by $\operatorname{sh} \pi u$ and integrate from 0 to K, obtaining

$$\frac{K^2}{2\pi} + \int_{u=0}^{K} \sum_{c=1}^{\infty} \frac{S(n, n; c)}{c} \frac{u}{\operatorname{ch} \pi u} \frac{4\pi n}{c} \int_{4\pi n/c}^{\infty} \{J_{2iu}(t) + J_{-2iu}(t)\} \frac{dt}{t}\, du$$

$$= \sum_{j=1}^{\infty} \frac{4\pi n|b_j(n)|^2}{\operatorname{ch} \pi\varkappa_j} \int_0^K \operatorname{sh} \pi u\, K(\varkappa_j, u)\, du + \int_0^K \int_{-\infty}^{\infty} \frac{\tau_{it}^2(n) \operatorname{sh} \pi u\, K(t, u)}{|\zeta(1+2it)|^2}\, dt\, du.$$

On the right we have an exact integral, since

$$\frac{\operatorname{sh} \pi u\, K(r,u)}{\operatorname{ch} \pi r} = \frac{\operatorname{sh} \pi u}{\operatorname{ch} \pi(r+u) \operatorname{ch} \pi(r-u)} = \frac{\operatorname{sh} \pi u}{\operatorname{ch}^2 \pi u + \operatorname{sh}^2 \pi r}$$

$$= \frac{1}{\pi \operatorname{sh} \pi r} \frac{d}{du} \tan^{-1} \frac{\operatorname{ch} \pi u}{\operatorname{sh} \pi r}$$

$$= \frac{1}{\pi \operatorname{sh} \pi r} \frac{d}{du} \{\tan^{-1} e^{\pi u - \pi r} - \tan^{-1} e^{-\pi u - \pi r}\},$$

integrating to an even function of K, which we can approximate as follows: for $1 \leqslant r \leqslant K - \log K$

$$\frac{1}{\pi \operatorname{sh} \pi r} \left\{\frac{\pi}{2} + O(e^{\pi r - \pi K})\right\} = \frac{1}{2 \operatorname{sh} \pi r} (1 + O(1/K^3));$$

and for $r \geqslant K + \log K$

$$\ll \frac{1}{\pi \operatorname{sh} \pi r} e^{\pi K - \pi r} \ll \frac{1}{K^3 \operatorname{sh} \pi r}.$$

The first term on the left is the main term; to estimate the second term we need one of Kuznetsov's careful estimates.

LEMMA. *For* $a > 0$,

$$\int_{u=0}^{K} \int_{x=a}^{\infty} \frac{u}{\operatorname{ch} \pi u} \{J_{2iu}(x) + J_{-2iu}(x)\} \frac{dx}{x} du \ll \min(1, 1/\sqrt{a}).$$

Proof. From Section 9 we see that

$$\frac{\pi}{4 \operatorname{ch} \pi u} (J_{2iu}(x) + J_{-2iu}(x))$$

$$= \int_0^T \sin(x \operatorname{ch} t) \cos 2tu\, dt + O\left(\frac{1+u}{xe^T}\right)$$

$$= \frac{\sin 2Tu}{2u} \sin(x \operatorname{ch} T) - \int_0^T \frac{x}{2u} \sin 2tu \cos(x \operatorname{ch} t) \operatorname{sh} t\, dt + O\left(\frac{1+u}{xe^T}\right).$$

Hence we have

$$\frac{\pi}{4}\int_a^b \frac{\{J_{2iu}(x)+J_{-2iu}(x)\}}{x \operatorname{ch} \pi u}\,dx$$

$$=\frac{\sin 2Tu}{2u}\int_a^b \frac{\sin (x \operatorname{ch} T)}{x}\,dx-\int_0^T \frac{\sin 2tu}{2tu}\cdot\frac{\{\sin (b \operatorname{ch} t)-\sin (a \operatorname{ch} t)\} \operatorname{sh} t}{\operatorname{ch} t}\,dt+$$

$$+O\left(\frac{1+u}{ae^T}\right),$$

and thus

$$\frac{\pi}{4}\int_{u=0}^K \int_{x=a}^b \frac{u}{\operatorname{ch} \pi u}\,\frac{\{J_{2iu}(x)+J_{-2iu}(x)\}}{x}\,dx$$

$$=\frac{(1-\cos 2TK)}{4T}\int_a^b \frac{\sin (x \operatorname{ch} T)}{x}\,dx-$$

$$-\int_0^T \frac{(1-\cos 2Kt)}{4t}\{\sin (b \operatorname{ch} t)-\sin (a \operatorname{ch} t)\} \operatorname{th} t\,dt+O\left(\frac{K+K^2}{ae^T}\right).$$

When we let $T\to\infty$, the first and third terms go out. Letting now $b\to\infty$, the term in $\sin (b \operatorname{ch} t)$ goes out on integration by parts, and the integral of the lemma is

$$\frac{1}{\pi}\int_0^\infty \frac{\operatorname{th} t}{t}(1-\cos 2Kt)\sin (a \operatorname{ch} t)\,dt.$$

As t increases $\operatorname{sh} t/t$ increases, and $\operatorname{th} t/t$ decreases. By the Second Mean Value Theorem, the integral is

$$\frac{1}{\pi}\max \frac{\operatorname{th} t}{t}\int_\xi^\infty (1-\cos 2Kt)\sin (a \operatorname{ch} t)\,dt$$

for some ξ, and for $a\geqslant 1/2$ we use van der Corput's second derivative test to estimate the integral as

$$\ll (\min a \operatorname{ch} t)^{-1/2}\ll 1/\sqrt{a}.$$

For $a \leqslant 1/2$ we treat the range $t \ll \log 1/a$ separately:

$$\int_0^{\log 1/a} \ll \int_0^{\log 1/a} \frac{a \operatorname{sh} t}{t}\, dt \ll a \log \frac{1}{a} \max \frac{\operatorname{sh} t}{t} \ll 1,$$

whilst as in the case $a \geqslant 1/2$, we have

$$\int_{\log 1/a}^{\infty} \ll \left(\max_{t \geqslant \log 1/a} \frac{\operatorname{th} t}{t}\right)\left(\min_{t \geqslant \log 1/a} a \operatorname{ch} t\right)^{-1/2} \ll 1/\log 1/a.$$

We see that

$$\int_{u=0}^{K} \sum_{c=1}^{\infty} \frac{S(n, n; c)}{c} \frac{u}{\operatorname{ch} \pi u} \frac{4\pi n}{c} \int_{4\pi n/c} \{J_{2iu}(t) + J_{-2iu}(t)\} \frac{dt}{t}\, du$$

$$\ll \sum_{c=1}^{\infty} \frac{n|S(n, n; c)|}{c^2} \min(1, \sqrt{c/n}).$$

Using Weil's estimate

$$|S(n, n; c)|^2 \leqslant d^2(c)(n, c)c,$$

we see that the sum is

$$\ll n \left\{\sum_{c=1}^{\infty} \frac{d(c)}{c^{3/2}} \min(1, \sqrt{c/n})\right\}^{1/2} \left\{\sum_{c=1}^{\infty} \frac{d(c)(n, c)}{c^{3/2}} \min(1, \sqrt{c/n})\right\}^{1/2}.$$

The first sum is $\ll \log^2 n/\sqrt{n}$, and the second is

$$\sum_{d|n} \frac{d(d)\varphi(d)}{d^{3/2}} \sum_{e=1}^{\infty} \frac{d(e)}{e^{3/2}} \min\left(1, \sqrt{\frac{e}{n/d}}\right) \ll \sum_{d|n} \frac{d(d)\varphi(d)}{d\sqrt{n}} \log^2 n,$$

and the original sum is

$$\ll (d_3(n))^{1/2} n^{1/2} \log^2 n.$$

The integral over t on the right is positive, so that changing K to $K + \log K$ we have the upper bound of the lemma. The asymptotic equality follows

from taking also $K-\log K$ for K and estimating the integral on the right non-trivially.

Further reading
KUBOTA
KUZNETSOV
ZAGIER

14. The Second Trace Formula

The Second Trace Formula corresponds to Perron's formula applied to the Kloosterman sum zeta function

$$Z_{m,n}(s) = \sum_1^\infty \frac{S(m, n; c)}{c^{2s}},$$

defined in the first instance for Re $s > 1$ but continuable to Re $s > 3/4$ from Weil's bound. We take

$$h(v) = \tfrac{1}{2} \sin \pi s \, \underline{|s-3/2+iv} \; \underline{|s-3/2-iv}$$

in the First Trace Formula. By Stieltjes Lemma (Corollary 2)

$$\int_{-\infty}^{\infty} v \operatorname{th} \pi v \, h(v) \, dv = \frac{\pi}{2} \frac{\underline{|s-1}}{\underline{|-s}}.$$

The main term on the left of the First Trace Formula is

$$\int_{-\infty}^{\infty} \frac{v h(v)}{\operatorname{ch} \pi v} J_{2iv}(\alpha) \, dv.$$

We take the integral to Im $v = -N$, N a large integer.

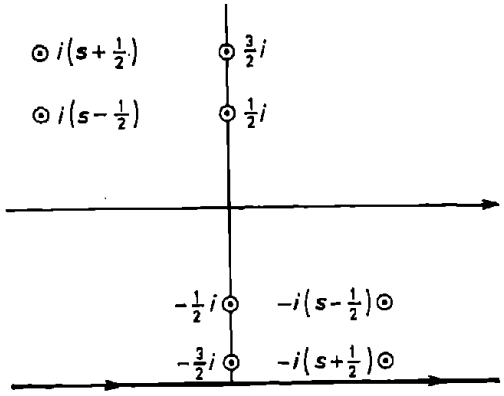

Fig. 33

When we move the contour $h(v)$ stays about the same size, except for poles at $v = -i(s-\frac{1}{2}+k)$, $k = 0, 1, 2, \ldots$ The ch πv factor is periodic with poles at $-(k-\frac{1}{2})i$, $k = 1, 2, \ldots$, and for large v

$$J_{2iv}(\alpha) \ll \frac{|\alpha^{2iv}|}{|\underline{2iv}|},$$

which is smaller by a factor about $\alpha^{2N}/\underline{2N}$. The integrand is tending to zero as the integer N tends to infinity, so that the integral along the real axis is $-2\pi i$ times the residues in the lower half plane. The residue at $v = -(k-1/2)i$ is

$$\frac{-(k-\frac{1}{2})i\,\underline{s+k-2}\,\underline{s-k-1}}{\pi i \sin \pi(k-1/2)} \frac{\sin \pi s}{2} J_{2k-1}(\alpha)$$

$$= \frac{(-1)^k(2k-1)}{4\pi} \sin \pi s\, \underline{s+k-2}\, \underline{s-k-1}\, J_{2k-1}(\alpha)$$

$$= \frac{2k-1}{4} \frac{\underline{s+k-2}}{\underline{k-s}} J_{2k-1}(\alpha),$$

using the functional equation for the factorial function

$$\underline{u}\,\underline{-u} = \pi u/\sin \pi u.$$

The residue at $v = -i(s-\frac{1}{2}+k)$ is

$$\frac{\sin \pi s}{2} \frac{\underline{2s-2+k}\ (-1)^k(-i)(s-\frac{1}{2}+k)}{\underline{k}\ (-i)\ \mathrm{ch}\left(-i\pi(s-\frac{1}{2}+k)\right)} J_{2s+2k-1}(\alpha)$$

$$= \frac{1}{4} \frac{\underline{2s+k-2}}{\underline{k}} (2s+2k-1) J_{2s+2k-1}(\alpha).$$

Putting back the factors $2i/\pi$ and $-2\pi i$ outside removes the factor 1/4. The sum of the residues at $-i(s-\frac{1}{2}+k)$ gives Neumann's series for $(\alpha/2)^{2s-1}$. For the other residues we have surprising result

$$\sum_c \frac{S(m, n; c)}{c} J_{2k-1}(4\pi\sqrt{mn/c^2}) = \frac{(-1)^k}{2\pi} \sqrt{\frac{m^{2k-1}}{n^{2k-1}}} \{c_n(G_{2k,m}) - \delta_{mn}\},$$

where $G_{2k,m}$ is the classical (holomorphic) Poincaré series of weight $2k$, and c_n denotes its nth Fourier coefficient. These terms appear in answer to Bruggemann's philosophy that all representations of PSL(2, R) over the cosets of the modular group PSL(2, Z) are equally important. To my mind

their presence hints that Kuznetsov's formulae are deeper than Selberg's Trace Formula. However in the sequel we shall not use this identification of the residues, but we transform the sum over Bessel functions yet again.

We have now shown that for $\operatorname{Re} s > 3/4$

$$(2\pi\sqrt{mn})^{2s-1} Z_{m,n}(s) +$$

$$+\frac{\delta_{mn}}{2\pi}\frac{\lfloor s-1}{\lfloor -s}+\sum_{k=1}^{\infty}\frac{(-1)^k}{2\pi}\sqrt{\frac{m^{2k-1}}{n^{2k-1}}}\{c_n(G_{2k,m})-\delta_{mn}\}(2k-1)\frac{\lfloor k+s-2}{\lfloor k-s}$$

$$=\sum_{j=1}^{\infty}\frac{4\sqrt{mn}\,b_j(m)b_j(n)}{2\operatorname{ch}\pi\varkappa_j}\sin\pi s\,\lfloor s-3/2+i\varkappa_j\,\lfloor s-3/2-i\varkappa_j+\zeta_{m,n}(s),$$

where $\zeta_{m,n}(s)$ is the integral over t, which will be considered next.

$$\zeta_{m,n}(s)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\tau_w(m)\tau_w(n)\sin\pi s\,\lfloor s-3/2+w\,\lfloor s-3/2-w}{\zeta(1+2w)\zeta(1-2w)}\,dw.$$

The integral makes sense for all s with $\operatorname{Re} s \neq 1/2-j$, where $j \geqslant 0$ is an integer (when there are poles on the contour), since the factorial functions fall off exponentially. However, if $\operatorname{Re} s < 1/2$, we must add the residues at pairs of poles to make the integral represent the same function.

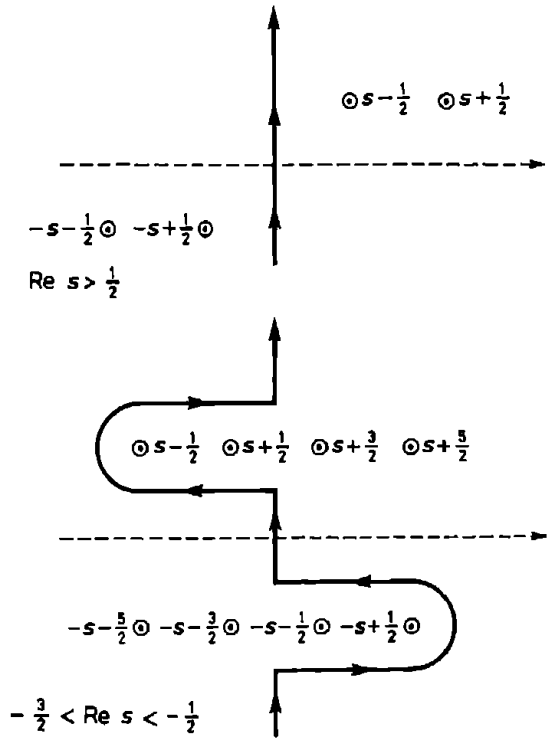

Fig. 34

Since the integrand is even in w, the residues are opposite, but corresponding poles are described in opposite senses. The residue of $\underline{|w+s-3/2}$ at $w = 1/2-s-k$ is $(-1)^k/\underline{|k}$, so that the whole integrand has residue

$$\tau_{k+s-1/2}(m)\tau_{k+s-1/2}(n)\frac{\sin \pi s \,\underline{|2s-2+k}}{\zeta(2-2s-2k)\zeta(2s+2k)}\cdot\frac{(-1)^k}{\underline{|k}}$$

$$= \frac{-\tau_{k+s-1/2}(m)\tau_{k+s-1/2}(n)}{\zeta(2-2s-2k)\zeta(2s+2k)}\cdot\frac{\pi}{2\cos \pi s \,\underline{|1-2s+k}\,\underline{|k}}$$

by the functional equation. Note that this residue only occurs when $\operatorname{Re} s < 1/2-k$, so that $\operatorname{Re}(2-2s-2k) > 1$. However $\zeta_{m,n}(s)$ does have a pole when $s = \varrho/2-k$, where ϱ is a zero of $\zeta(s)$. For $-K-1/2 < \operatorname{Re} s < -K+1/2$ we have

$$\zeta_{m,n}(s) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\tau_w(m)\tau_w(n)\sin \pi s \frac{\underline{|s-3/2+w}\,\underline{|s-3/2-w}}{\zeta(1+2w)\zeta(1-2w)}\,dw -$$

$$-\sum_{k=0}^{K}\frac{\pi}{\cos \pi s\,\underline{|k}}\cdot\frac{\tau_{s+k-1/2}(m)\tau_{s+k-1/2}(n)}{\zeta(2-2s-2k)\zeta(2s+2k)\,\underline{|1-2s+k}};$$

when $\operatorname{Re} s+1/2$ is an integer, we must indent the contour appropriately.

Our conclusion is that $\zeta_{m,n}(s)$ is a meromorphic function of s, so that $Z_{m,n}(s)$ is expressed as a convergent sum of meromorphic functions, with poles at $-k$, $1/2+i\varkappa_j-k$ and $\varrho/2-k$, where $j \geqslant 1$, $k \geqslant 0$ are integers and ϱ is a zero of the zeta function.

Let $g(x)$ be a piecewise differentiable continuous weight function, non zero on $a \leqslant x \leqslant b$, a finite interval. With a slight change of notation from Section 8, we write the Mellin transform as

$$G(s) = -\int_a^b \frac{x^{2s}}{s}\,g'(x)\,dx = \int_0^\infty 2x^{2s-1}\,g(x)\,dx.$$

On the line L, $\operatorname{Re} s = 5/4$ we have the inverse formula

$$\int_L \frac{G(s)\,ds}{y^{2s}} = -\int_a^b\int_L \left(\frac{x}{y}\right)^{2s} g'(x)\,\frac{ds}{s}\,dx = -\int_y^b g'(x)\,dx = g(y).$$

Since

$$G(s) \ll (a^{2\sigma}+b^{2\sigma})/|s|^2,$$

the integral

$$\hat{g}(v) = \frac{1}{2\pi i} \int_L \tfrac{1}{2} G(s) \sin \pi s \, \underline{\lfloor s-3/2+iv} \, \underline{\lfloor s-3/2-iv} \, ds$$

converges absolutely. As in Section 9, we may change the line of integration to a loop M from $-\infty$ round all the poles of the functions. By absolute convergence of the double integral

$$\hat{g}(v) = \frac{1}{2\pi i} \int_a^b \int_M x^{2s-1} g(x) \sin \pi s \, \underline{\lfloor s-3/2+iv} \, \underline{\lfloor s-3/2-iv} \, dx\, ds$$

$$= \frac{\pi i}{2 \operatorname{sh} \pi v} \int_{x=a}^{b} \{J_{2iv}(2/x) - J_{-2iv}(2/x)\} g(x)\, dx.$$

We also have

$$\hat{g}(v) \ll v^{-5/2} \quad \text{as } v \to +\infty, \quad \text{in } |\operatorname{Im} v| \leqslant \pi/2;$$

this can be seen either from the integral over L or from integrating the Bessel functions term by term as power series twice. Similarly we have

$$\frac{1}{2\pi i} \int_L G(s) \frac{\underline{\lfloor s-1+u}}{\underline{\lfloor u-s}} \, ds = \int_a^b J_{2u}\left(\frac{2}{x}\right) \frac{g(x)}{x} \, dx$$

and

$$\frac{1}{2\pi i} \int_L G(s) \frac{\underline{\lfloor s-2+u}}{\underline{\lfloor u-s}} \, ds = \int_a^b J_{2u+1}\left(\frac{2}{x}\right) g(x)\, dx.$$

We substitute $\hat{g}(v)$ as the $h(v)$ in the First Trace Formula, to obtain Kuznetsov's main result, the Second Trace Formula.

THEOREM. *Let $g(x)$ be continuous of compact support and piecewise differentiable, and $\hat{g}(v)$ be the Kuznetsov transform defined above. Then*

$$\frac{1}{2\pi\sqrt{mn}} \sum_{c=1}^{\infty} g\left(\frac{c}{2\pi\sqrt{mn}}\right) S(m, n; c) + \sum_{k=1}^{\infty} \frac{(-1)^k}{2\pi} \sqrt{\frac{m^{2k-1}}{n^{2k-1}}} \{c_n(G_{2k,m}) - \delta_{mn}\} \times$$

$$\times (2k-1) \int J_{2k-1}\left(\frac{2}{x}\right) g(x)\, dx + \delta_{mn} \int J_0\left(\frac{2}{x}\right) \frac{g(x)}{x} \, dx$$

$$= \sum_{j=1}^{\infty} \frac{4\sqrt{mn}\, \overline{b_j(m)}\, b_j(n)\, \hat{g}(\varkappa_j)}{\operatorname{ch} \pi\varkappa_j} + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\tau_{it}(m)\, \tau_{it}(n)\, \hat{g}(t)}{\zeta(1+2it)\, \zeta(1-2it)} \, dt.$$

Proof. When we apply the First Trace Formula to the right-hand side, we obtain

$$\frac{1}{2\pi i}\int_L G(s)(2\pi\sqrt{mn})^{2s-1} Z_{m,n}(s)\,ds + \frac{\delta_{mn}}{2\pi i}\int_L \frac{G(s)\,\lfloor s-1}{\lfloor -s}\,ds +$$

$$+ \sum_{k=1}^{\infty} \frac{(-1)^k}{2\pi}\sqrt{\frac{m^{2k-1}}{n^{2k-1}}}\{c_n(G_{2k,m})-\delta_{mn}\}\frac{(2k-1)}{2\pi i}\int_L \frac{G(s)\,\lfloor k+s-2}{\lfloor k-s}\,ds.$$

The integral involving $Z_{m,n}(s)$ is an ordinary Perron integral which gives the weighted sum of Kloosterman sums term by term. The other two integrals have been noted above.

Our notation for the transform here differs from Kuznetsov and Deshouillers–Iwaniec. Kuznetsov writes $g(x)$ as $u\varphi(u)$, where $u = 2/x$. For applications it seems more natural to take x as the independent variable, not its inverse.

Finally we indicate how to treat the sum over coefficients of classical Poincaré series, by a typically ingenious argument of Kuznetsov. Multiplying the recurrences for Bessel functions, we find that

$$(J_{s+1}(x)+J_{s-1}(x))(J_{s+1}(y)-J_{s-1}(y)) = \frac{-4s}{x}J_s(x)J_s'(y).$$

Interchanging x and y and adding, we have

$$J_{s+1}(x)J_{s+1}(y)-J_{s-1}(x)J_{s-1}(y) = -\frac{2s}{x}J_s(x)J_s'(y)-\frac{2s}{y}J_s(y)J_s'(x),$$

so that

$$\frac{-2s}{xy}\frac{d}{dt}J_s(xt)J_s(yt) = tJ_{s+1}(xt)J_{s+1}(yt)-tJ_{s-1}(xt)J_{s-1}(yt).$$

For $\operatorname{Re} s > 0$ the Bessel functions can be integrated from 0 to 1 to give

$$\frac{2s}{xy}J_s(x)J_s(y) = \int_0^1 tJ_{s-1}(xt)J_{s-1}(yt)\,dt - \int_0^1 tJ_{s+1}(xt)J_{s+1}(yt)\,dt.$$

Adding these equations, we have

$$\frac{2}{xy}\sum_{k=1}^{K}(2k-1)J_{2k-1}(x)J_{2k-1}(y) = \int_0^1 tJ_0(xt)J_0(yt)\,dt - \int_0^1 tJ_{2K}(xt)J_{2K}(yt)\,dt.$$

Now as $K \to \infty$, $J_{2K}(u)$ tends to zero uniformly for $u \leqslant \max(x, y)$ and we have

$$\sum_1^\infty (2k-1) J_{2k-1}(x) J_{2k-1}(y) = \tfrac{1}{2} xy \int_0^1 t J_0(xt) J_0(yt)\, dt.$$

Recall that

$$\frac{(-1)^k}{2\pi} \sqrt{\frac{m^{2k-1}}{n^{2k-1}}} \{c_n(G_{2k,m}) - \delta_{mn}\} = \sum_c \frac{S(m, n; c)}{c} J_{2k-1}(\alpha(c)).$$

The sum over Poincaré series in the Second Trace Formula can now be written as

$$\sum_k \frac{(-1)^k}{2} \sqrt{\frac{m^{2k-1}}{n^{2k-1}}} \{c_n(G_{2k,m}) - \delta_{mn}\}(2k-1) \int_a^b J_{2k-1}\left(\frac{2}{x}\right) g(x)\, dx$$

$$= \sum_c \frac{4\pi\sqrt{mn}}{c} \cdot \frac{S(m, n; c)}{c} \int_a^b \frac{g(x)}{x} \int_{t=0}^1 t J_0\left(\frac{2t}{x}\right) J_0\left(\frac{4\pi t\sqrt{mn}}{c}\right) dt\, dx,$$

a more convenient form for order of magnitude estimates.

Further reading
KUZNETSOV
RANKIN

15. Linnik's Sum

Linnik was the first to conjecture that a sum of Kloosterman sums $S(m, n; c)$ with fixed m and n and varying c should behave like a sum of random variables with root mean square $c^{1/2+\varepsilon}$. Kuznetsov's method allows us to estimate smoothly weighted sums very well. We give only the simplest bound here. Deshouillers and Iwaniec have many multiple averages, often with congruence conditions that necessitate working in congruence subgroups of the full modular group.

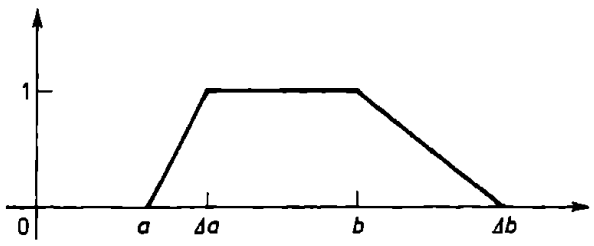

Fig. 35

We take

$$g(x) = \frac{1}{\Delta - 1} \int_{\max(1, x/b)}^{\min(\Delta, x/a)} dy,$$

where $\Delta \leqslant 2$ and $a\Delta \leqslant b$. We have for any integrable function $f(x)$

$$\int_0^\infty f(x)\, g(x)\, dx = \int_{x=0}^{\infty} \frac{1}{\Delta - 1} \int_{y=\max(1, x/b)}^{\min(\Delta, x/a)} f(x)\, dy\, dx$$

$$= \int_{y=1}^{\Delta} \int_{x=ay}^{by} \frac{f(x)}{\Delta - 1}\, dx\, dy = \int_{y=1}^{\Delta} \int_{z=a}^{b} \frac{f(yz)}{\Delta - 1}\, y\, dz\, dy.$$

Hence the Kuznetsov transform is

$$\hat{g}(v) = \frac{\pi i}{2 \operatorname{sh} \pi v} \int_{x=a}^{b} \left\{ J_{2iv}\left(\frac{2}{x}\right) - J_{-2iv}\left(\frac{2}{x}\right) \right\} g(x)\, dx$$

$$= \frac{\pi i}{2 \operatorname{sh} \pi v} \int_{y=1}^{\Delta} \int_{z=a}^{b} \frac{2i}{\Delta - 1} \operatorname{Im} J_{2iv}\left(\frac{2}{yz}\right) y\, dz\, dy.$$

If either $a \gg 1$ or $a \gg 1/v$ we can write the Bessel function as a power series

$$J_{2iv}\left(\frac{2}{yz}\right) = \frac{1}{\lfloor 2iv} \left(\frac{1}{yz}\right)^{2iv} - \frac{1}{\lfloor 1 \lfloor 2iv+1} \left(\frac{1}{yz}\right)^{2iv+2} + \ldots$$

in which the first term dominates, and

$$\hat{g}(v) \ll \frac{b-a}{\operatorname{sh} \pi v} \max_{a \leqslant \xi \leqslant \Delta b} \left| \operatorname{Im} \frac{1}{\lfloor 2iv} \left(\frac{1}{\xi}\right)^{2iv} + \ldots \right|$$

$$\ll \begin{cases} (b-a)(1 + |\log a| + |\log b|) & \text{if } v \leqslant 1, \\ (b-a)/\sqrt{v} & \text{if } v \geqslant 1. \end{cases}$$

We can also perform one or both of the integrations, giving

$$\hat{g}(v) \ll \frac{1}{\operatorname{sh} \pi v} \max_{y} \left| \operatorname{Im} \frac{b}{2iv - 1} \left(\frac{1}{by}\right)^{2iv} \frac{1}{\lfloor iv} + \ldots \right| \ll b/v^{3/2}$$

or

$$\hat{g}(v) \ll \frac{1}{\operatorname{sh} \pi v} \frac{1}{\Delta - 1} \max_{\xi} \left| \operatorname{Im} \frac{\xi^{1-2iv}}{(2iv-1)(2iv-2)} \frac{1}{\underline{iv}} + \ldots \right| \ll \frac{b}{(\Delta-1)v^{5/2}}.$$

If the condition on a is not satisfied, we must have $0 \leqslant v \leqslant 1$, and then

$$\frac{\pi i}{4 \operatorname{sh} \pi v} \left\{ J_{2iv}\left(\frac{2}{yz}\right) - J_{-2iv}\left(\frac{2}{yz}\right) \right\} = \int_0^\infty \cos\left(\frac{2 \operatorname{ch} t}{yz}\right) \cos(2tv)\, dt \ll \sqrt{yz}$$

by van der Corput's second derivative test, so that

$$\hat{g}(v) \ll \int_a^b \sqrt{x}\, g(x)\, dx \ll (b-a)\sqrt{b}.$$

If b is large we break the range of integration: for $0 \leqslant v \leqslant 1$, $b \geqslant e$

$$\int_0^{\log b} + \int_{\log b}^\infty \ll \log b + 1 \quad \text{and} \quad \hat{g}(v) \ll (b-a) \log b.$$

We substitute $g(x)$ in the Second Trace Formula, and use Cauchy's inequality

$$\left| \sum_j \frac{4\sqrt{mn}\, \overline{b_j(m)}\, b_j(n)\, \hat{g}(\varkappa_j)}{\operatorname{ch} \pi \varkappa_j} \right|^2 \leqslant \left\{ \sum \frac{4n |b_j(m)|^2 |\hat{g}(\varkappa_j)|}{\operatorname{ch} \pi \varkappa_j} \right\} \left\{ \sum \frac{4n |b_j(n)|^2 |\hat{g}(\varkappa_j)|}{\operatorname{ch} \pi \varkappa_j} \right\}.$$

We now use the estimate for the sum of squares of Fourier coefficients in Section 13.

$$\sum_1^\infty \frac{4n |b_j(n)|^2 |\hat{g}(\varkappa_j)|}{\operatorname{ch} \pi \varkappa_j} \ll \int_1^\infty (v^2 + d_3(n)\sqrt{n} \log^2 n)\, d|\hat{g}(v)|$$

$$\ll \int_1^\infty v |\hat{g}(v)|\, dv + (d_3(n))^{1/2} \sqrt{n} \log^2 n \max_{v \geqslant 1} |\hat{g}(v)|.$$

Using our bounds for the transform $\hat{g}(v)$, we estimate the sum as

$$\ll \int_1^K \frac{vb}{v^{3/2}}\, dv + \int_K^\infty \frac{vb}{(\Delta-1)v^{5/2}}\, dv + (b-a)(d_3(n))^{1/2} \sqrt{n} \log^2 n$$

$$\ll b\sqrt{K}+\frac{b}{(\Delta-1)\sqrt{K}}+(b-a)(d_3(n))^{1/2}\sqrt{n}\log^2 n$$

$$\ll \frac{b}{(\Delta-1)}+(b-a)(d_3(n))^{1/2}\sqrt{n}\log^2 n,$$

when we optimise the choice of K.

More easily, for $a \gg 1$

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\tau_{it}(m)\tau_{it}(n)\hat{g}(t)}{\zeta(1+2it)\zeta(1-2it)}\,dt$$

$$\ll \int_0^2 d(m)d(n)t^2(b-a)(\log b+1)\,dt+\int_2^{\infty} d(m)d(n)\frac{\log^2 t}{t^{3/2}}\,b\,dt$$

$$\ll d(m)d(n)\{b+(b-a)\log b\}.$$

The sum over classical Poincaré series has already been transformed; it may now be bounded

$$\sum_c \frac{4\pi\sqrt{mn}}{c}\frac{S(m,n;c)}{c}\int_{x=a}^{b}\frac{g(x)}{x}\int_{t=0}^{1} tJ_0\left(\frac{2t}{x}\right)J_0\left(\frac{4\pi t\sqrt{mn}}{c}\right)dt\,dx$$

$$\ll \sum_c \sqrt{mn}\frac{|S(m,n;c)|}{c^2}\int_{x=a}^{b}\frac{|g(x)|}{x}\,dx\,\min\left(1,\frac{c^{1/2}}{(mn)^{1/4}}\right),$$

where we have used

$$J_0(y) \ll \min(1, y^{1/2}).$$

The whole expression is now

$$\ll \log\frac{b}{a}(d_3((m,n)))^{1/2}(mn)^{1/4}\log^2 mn,$$

by the calculation used to prove the Mean Value Lemma. If $b < 1$ we should estimate both Bessel functions by the square root of the variable. Finally

$$\int_a^b J_0\left(\frac{2}{x}\right)\frac{g(x)}{x}\,dx \ll \int_a^b \frac{g(x)}{x}\,dx \ll \log\frac{b}{a} \ll \frac{b-a}{a}.$$

Writing $M = 2\pi a\sqrt{mn}$, $N = 2\pi b\sqrt{mn}$, for $a \gg 1$

$$\sum_c g\left(\frac{c}{2\pi\sqrt{mn}}\right) S(m, n; c)$$

$$\ll \frac{N}{\sqrt{\Delta-1}} \left\{1+\sqrt{\Delta-1}\left(1-\frac{M}{N}\right) d(n)\sqrt{n}\log^2 n\right\}^{1/2} \times$$

$$\times \left\{1+\sqrt{\Delta-1}\left(1-\frac{M}{N}\right) d(m)\sqrt{m}\log^2 m\right\}^{1/2}.$$

There is no point in considering $M < \sqrt{mn}$ here, as the term in $(N-M)(mn)^{1/4}$ is larger than one would obtain by summing Weil's bound. Deshouillers and Iwaniec get nontrivial results in the range $N \ll \sqrt{mn}$ by averaging over other parameters. However without extra averaging, we get the main term $\ll N$ of the Linnik conjecture if m and n are bounded, and Δ may be bounded away from zero.

For an unweighted sum we must remove the tapering. This introduces an extra error term

$$\ll \left(\sum_M^{M\Delta} + \sum_N^{N\Delta}\right) |S(m, n; c)|.$$

We use Weil's bound in the form

$$|S(m, n; q)|^2 \leqslant d^2(q)(2m, 2n, q)$$

so that

$$|S(m, n; q)| \ll d(q)\sqrt{q} = \sum_{d|q} f(d)$$

where

$$\zeta(s)\sum_1^\infty \frac{f(d)}{d^s} = \zeta^2\left(\frac{s}{2}\right).$$

The new error term is now

$$\ll \sqrt{N} \sum_{d|(m,n)} f(d)\left\{\frac{N(\Delta-1)}{d}+1\right\}$$

$$\ll (\Delta-1)N^{3/2} \sum_{d|(m,n)} \frac{f(d)}{d} + d((m, n))\sqrt{mnN}.$$

Here we have the estimate

$$\sum_{d|(m,n)} \frac{f(d)}{d} \leqslant \prod_{p|(m,n)} \left(1-\frac{1}{p}\right)\left(1-\frac{1}{\sqrt{p}}\right)^2 \leqslant v^2((m, n))$$

where we define

$$v(m) = \prod_{p|m} \left(1 - \frac{1}{\sqrt{p}}\right)^{-1}.$$

Choosing

$$\Delta - 1 = N^{-1/3} \left(v((m, n))\right)^{-4/3}$$

we obtain Kuznetsov's unweighted bound (which he states slightly differently, without explicit dependence on m and n).

THEOREM. *For* $\sqrt{mn} \ll M \leqslant N$

$$\sum_{q=M}^{N} S(m, n; q) \ll N^{7/6} \left(v((m, n))\right)^{2/3} +$$
$$+ N^{13/12} \left(v((m, n))\right)^{1/3} \{d(m)\sqrt{m} \log m + d(n)\sqrt{n} \log n\} +$$
$$+ N d(m) d(n) (mn)^{1/4} \log^2 mn,$$

where $v(m)$ *was defined above.*

We get $N^{7/6}$ in the leading term because we estimated $|\hat{g}(v)|$. If the Fourier coefficients $b_j(n)$ were sufficiently regular, we could make some use of the oscillation of $\hat{g}(v)$. This seems out of reach at present.

Further reading
DESHOUILLERS and IWANIEC
KUZNETSOV

16. Kloosterman sums with opposite signs

To prove the corresponding formulae for $S(-m, n; c)$ we consider instead

$$(U_m(z, s), \overline{U_n(z, w)}) = \int_{y=0}^{\infty} e^{-2\pi my} a_{-m} y^{-2} dy$$

in the notation of Section 12. The final formulae have I Bessel functions in place of J functions, but the intermediate steps are quite different. The expression for the Kloosterman sum zeta function as a sum of meromorphic functions fails to converge absolutely, so we follow Kuznetsov's treatment here. The holomorphic Poincaré series do not arise here, as they have no terms in $e(mz)$ with $m < 0$.

For $1 < \operatorname{Re} w < \operatorname{Re} s$ the inner product is expressed as

$$\int_{y=0}^{\infty} y^s e^{-\pi my} \int_{u=-\infty}^{\infty} \exp\left(-2\pi imty - \frac{2\pi n}{c^2 y(1+it)}\right) \frac{1}{(t^2+1)^w} dt\, dy,$$

the only change being the sign of im. However in

$$\mathrm{Kh}_s(x) = \frac{1}{\pi}\int_0^\infty \exp\left(-\frac{x}{2}\left(u+\frac{1}{u}\right)\right)\frac{du}{u^{s+1}},$$

we must put

$$x = \alpha(c) = 4\pi\sqrt{mn/c^2}, \qquad u = (1+it)\,y\sqrt{\frac{mc^2}{n}}$$

so that the subsequent variable change $v = (1-it)^{1/2}(1+it)^{-1/2}$ gives

$$\left(U_m(z, s), \overline{U_n(z, w)}\right)$$

$$= \sum_{c=1}^\infty \frac{S(-m, n; c)}{c^{s+w}}\left(\frac{n}{m}\right)^{(s-w)/2} \pi i \mathrm{Kh}_{s-w}(\alpha(c)) \int_C \left\{\tfrac{1}{2}\left(v+\frac{1}{v}\right)\right\}^{s+w-2} v^{s-w}\frac{dv}{v}$$

on the same contour C as before. The continuation to $\mathrm{Re}\, s > 3/4$, $\mathrm{Re}\, w > 3/4$ again follows. When we put $s = 1+iu$, $w = 1-iu$, the integral over C, half the unit circle from $-i$ to i, becomes

$$\int_C v^{2iu-1}\,dv = \frac{i \operatorname{sh} \pi u}{u}.$$

This leads to a slightly simpler Basic Lemma.

Basic Lemma. *For* $-1/4 < \mathrm{Im}\, u < 1/4$

$$-\sum_{c=1}^\infty \frac{S(-m, n; c)}{c}\alpha(c)\,\mathrm{Kh}_{2iu}(\alpha(c))$$

$$= \sum_{j=1}^\infty \frac{4\pi\sqrt{mn}\,b_j(m)b_j(n)}{\operatorname{ch} \pi\varkappa_j} K(\varkappa_j, u) + \int_{-\infty}^\infty \frac{\tau_{it}(m)\tau_{it}(n)K(t, u)}{\zeta(1+it)\zeta(1-it)}\,dt.$$

Recall from Section 11 that Maass wave forms are either pure cosine series or pure sine series. For cosine series $b_j(-m) = b_j(m)$ is real, for sine series $b_j(-m) = -b_j(m)$, which is pure imaginary. The Eisenstein series is a cosine series. Hence the $b_j(m)$ in the Basic Lemma is 'really' $\overline{b_j}(-m)$.

The deduction of the First Trace Formula follows very similar lines. We multiply the Basic Lemma by

$$\tfrac{1}{2}\operatorname{ch} \pi u\,\{h(u+i/2)+h(u-i/2)\}$$

and integrate over u. The Macdonald function $\mathrm{Kh}_{2iu}(\alpha)$ needs to be ex-

panded into I functions, giving

$$\int_{-\infty}^{\infty} \frac{\operatorname{ch} \pi u}{2} \cdot \frac{1}{\sin 2i\pi u} \left(I_{-2iu}(\alpha) - I_{2iu}(\alpha)\right) \{h(u+i/2) + h(u-i/2)\}\, du$$

$$= \frac{i}{2} \int_{-\infty}^{\infty} \frac{1}{\operatorname{sh} \pi u} \{h(u+i/2) + h(u-i/2)\}\, I_{2iu}(\alpha)\, du$$

since $h(u) = h(-u)$. We move each term to the real axis as before, getting

$$\frac{i}{2} \int_{-\infty}^{\infty} \frac{h(v) I_{2iv+1}(\alpha)}{-i \operatorname{ch} \pi v}\, dv + \frac{i}{2} \int_{-\infty}^{\infty} \frac{h(v) I_{2iv-1}(\alpha)}{i \operatorname{ch} \pi v}\, dv = \int_{-\infty}^{\infty} \frac{4iv}{\alpha} \frac{I_{2iv}(\alpha) h(v)}{2 \operatorname{ch} \pi v}\, dv$$

by the recurrence relation

$$I_{s-1}(\alpha) - I_{s+1}(\alpha) = \frac{2s}{\alpha} I_s(\alpha).$$

The convergence of the sum over c (so that $\alpha(c) \to 0$) follows as before, and we obtain the First Trace Formula.

THEOREM. *Let $h(z)$ be regular for $|y| \leqslant 1/2$ and tend to zero sufficiently fast for $|x| \to +\infty$. Then*

$$-2i \sum_{c=1}^{\infty} \frac{S(-m, n; c)}{c} \int_{-\infty}^{\infty} \frac{v h(v)}{\operatorname{ch} \pi v} I_{2iv}(4\pi \sqrt{mn}/c^2)\, dv$$

$$= \sum_{j=1}^{\infty} \frac{4\pi \sqrt{mn}\, b_j(m) b_j(n) h(\varkappa_j)}{\operatorname{ch} \pi \varkappa_j} + \int_{-\infty}^{\infty} \frac{\tau_{it}(m) \tau_{it}(n) h(t)}{\zeta(1+2it) \zeta(1-2it)}\, dt.$$

The convergence condition is the same as in Section 13. To obtain the Kloosterman sum zeta function, we should take

$$h(v) = \operatorname{ch} \pi v \, \underline{|s-3/2+iv} \, \underline{|s-3/2-iv},$$

which is small enough at infinity only for $\operatorname{Re} s < 0$. To get round this difficulty we must deduce the Second Trace Formula directly from the first. What follows is essentially the integral transform of Kontorovich and Lebedev. Suppose that $4T$ is an odd integer and $a \leqslant x \leqslant y \leqslant b$. Then

$$\int_{-iT}^{iT} 2\pi s\, I_{2s}\left(\frac{2}{y}\right) \operatorname{Kh}_{2s}\left(\frac{2}{x}\right) ds = \int_{-iT}^{iT} \frac{2\pi s}{\sin 2\pi s} I_{2s}\left(\frac{2}{y}\right) \left\{I_{-2s}\left(\frac{2}{x}\right) - I_{2s}\left(\frac{2}{x}\right)\right\} ds.$$

We move the contour to C, made up of
C_1: line segment $-iT$ to $T-iT$,
C_2: line segment $T-iT$ to $T+iT$,
C_3: line segment $T+iT$ to iT.
When τ is large we have

$$I_{2s}\left(\frac{2}{y}\right) \ll \begin{cases} e^{\pi T}/y^{2\sigma}\, T^{2\sigma+1} & \text{on } C_1 \text{ and } C_3, \\ e^{\pi T}/y^{2\sigma}(1+|t|)^{2\sigma+1} & \text{on } C_2. \end{cases}$$

The term in $I_{2s}I_{2s}$ integrates to $\ll 1/\log T$. We expand the term in $I_{2s}I_{-2s}$ as power series.

$$\begin{aligned} \frac{2\pi s}{\sin 2\pi s} I_{2s}\left(\frac{2}{y}\right) I_{-2s}\left(\frac{2}{x}\right) &= \underline{\lfloor 2s}\, \underline{\lfloor -2s}\, I_{2s}\left(\frac{2}{y}\right) I_{-2s}\left(\frac{2}{x}\right) \\ &= \left(\frac{x}{y}\right)^{2s} \left\{1+\frac{1}{\underline{\lfloor 1}\,(1+2s)y^2}+\frac{1}{\underline{\lfloor 2}\,(1+2s)(2+2s)y^4}+\dots\right\} \times \\ &\qquad \times\left\{1+\frac{1}{\underline{\lfloor 1}\,(1-2s)x^2}+\frac{1}{\underline{\lfloor 2}\,(1-2s)(2-2s)x^4}+\dots\right\} \\ &= \left(\frac{x}{y}\right)^{2s} \left\{1+\frac{2s}{1-4s^2}\left(\frac{1}{x^2}-\frac{1}{y^2}\right)+O\left(\frac{1}{|s|^2}\right)\right\}. \end{aligned}$$

The error term and the second term integrate to $O(1/T)$ uniformly in $a \leqslant x \leqslant y$. For $a \leqslant y \leqslant x \leqslant b$ we argue similarly, moving the contour to the left instead. Hence uniformly in $a \leqslant x \leqslant b$, $a \leqslant y \leqslant b$

$$\int_{-iT}^{iT} 2\pi s\, I_{2s}\left(\frac{2}{y}\right) \mathrm{Kh}_{2s}\left(\frac{2}{x}\right) ds = \int_{-iT}^{iT} \left(\frac{x}{y}\right)^{2s} ds + O\left(\frac{1}{\log T}\right).$$

Now let $g(x)$, $G(s)$ be the Mellin transform pair used in Section 14. Then

$$\begin{aligned} &\int_{-iT}^{iT} 2\pi s\, I_{2s}\left(\frac{2}{y}\right) \int_{x=a}^{b} g(x)\, \mathrm{Kh}_{2s}\left(\frac{2}{x}\right) dx\, ds \\ &= \int_{-iT}^{iT}\int_{a}^{b} \left(\frac{x}{y}\right)^{2s} g(x)\, dx\, dw + O\left(\frac{1}{\log T}\right) = \int_{-iT}^{iT} \frac{G(s+1/2)}{y^{2s}}\, ds + O\left(\frac{1}{\log T}\right), \end{aligned}$$

which tends to $2\pi i y g(y)$ as $T \to \infty$, and we have Kontorovich and Lebedev's

theorem in the single integral form

$$\int_{v=-\infty}^{\infty} iv\, I_{2iv}\left(\frac{2}{y}\right) \int_a^b g(x)\, \mathbf{K}h_{2iv}\left(\frac{2}{x}\right) dx\, dv = yg(y).$$

Kuznetsov deduces the Second Trace Formula in this manner in the case of like sign. There are now poles of the integrand $(J_{-s}-J_s)/\sin \pi s$ at $s = 1, 3, 5$ since $I_{-n} = I_n$ for all integers n, but $J_{-n} = (-1)^n J_n$. The single integral formula for the Kuznetsov transform

$$yg(y) = \sum_1^{\infty} (2k-1) J_{2k-1}\left(\frac{2}{y}\right) \int_a^b J_{2k-1}\left(\frac{2}{x}\right) g(x)\, dx +$$

$$+ \int_{-\infty}^{\infty} \frac{iv}{\operatorname{sh} 2\pi v} J_{2iv}\left(\frac{2}{y}\right) \int_a^b \left\{ J_{2iv}\left(\frac{2}{x}\right) - J_{-2iv}\left(\frac{2}{x}\right) \right\} g(x)\, dx\, dv$$

is proved in the same way.

To prove the Second Trace Formula, we put

$$\check{g}(v) = \frac{\operatorname{ch} \pi v}{2\pi i} \int_L \tfrac{1}{2} G(s) \underline{\lfloor s-3/2+iv} \underline{\lfloor s-3/2-iv}\, ds.$$

Again $\check{g}(v) \ll v^{-5/2}$ as $\operatorname{Re} v \to \infty$, $|\operatorname{Im} v| \leqslant 1/2$ and $\check{g}(v) \operatorname{sh} \pi v$ is regular for all v. The contour integral converges even when we replace L by a loop M from $-\infty - iT$ to $-\infty + iT$ passing all poles to the left. We can now write

$$g(v) = \frac{\operatorname{ch} \pi v}{2\pi i} \int_{x=a}^{b} \int_M \tfrac{1}{2} x^{2s-1} g(x) \underline{\lfloor s-3/2+iv} \underline{\lfloor s-3/2-iv}\, dx\, ds$$

$$= \frac{\pi}{2} \operatorname{ch} \pi v \int_a^b \mathbf{K}h_{2iv}\left(\frac{2}{x}\right) g(x)\, dx,$$

using a Bessel function integral from Section 9, inside an absolutely convergent double integral.

We can now deduce the Second Trace Formula.

THEOREM. *Let $g(x)$ be continuous of compact support and piecewise differentiable, and let $\check{g}(v)$ be its Kontorovich–Lebedev transform as defined*

above. Then

$$-\frac{1}{2\pi\sqrt{mn}}\sum_{c=1}^{\infty} g\left(\frac{c}{2\pi\sqrt{mn}}\right)S(-m, n; c)$$

$$= \sum_{j=1}^{\infty} \frac{4\sqrt{mn}\, b_j(m)\, b_j(n)\, \check{g}(\varkappa_j)}{\operatorname{ch} \pi\varkappa_j} + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\tau_{it}(m)\tau_{it}(n)\,\check{g}(t)\,dt}{\zeta(1+2it)\,\zeta(1-2it)}.$$

The bounds and estimates are the same as for $\hat{g}(v)$, since $\check{g}(v)$ is the I-Bessel function analogue of $\hat{g}(v)$. For small b there is a better bound since

$$|\mathrm{Kh}_{2iv}(2/b)| \leqslant \mathrm{Kh}_0(2/b) \ll \exp(-2/b).$$

We find that

$$\check{g}(v) \ll \begin{cases} (b-a)\exp(-2/b) & \text{for } b \leqslant 2, \\ (b-a)\log b & \text{for } b \geqslant 2, \end{cases}$$

which we can use for small v. However this improvement does not extend to the final bound for Linnik sums, and we get the same bound as in Section 15 for the sum

$$\sum_{q=M}^{N} S(-m, n; q).$$

The case of opposite signs appears in Kuznetsov's preprint, but not in the expanded account in Sbornik.

Further reading
KUZNETSOV

References

[1] O. Atkin and J. Lehner, *Hecke operators on $\Gamma^0(m)$*, Math. Ann. 185 (1970), 134–160.
[2] C. Bandle, *Isoperimetric inequalities and applications*, London 1980.
[3] R. W. Bruggeman, *Fourier coefficients of automorphic forms*, Springer Lecture Notes, 865, 1981.
[4] –, *Automorphic forms*, these Proceedings, pp. 31–74.
[5] –, *Fourier coefficients of cusp forms*, Invent. Math. 45 (1978), 1–78.
[6] P. Buser, *Riemannsche Flächen mit Eigenwerten in* (0, 1/4), Comment. Math. Helv. 52 (1977), 25–34.
[7] –, *Riemannsche Flächen mit grosser Kragenweite*, ibid. 53 (1978), 395–407.
[8] –, *On Cheeger's inequality $\lambda_1 \geqslant h^2/4$*, AMS Proc. Symposia Pure Maths. 36 (1980), 29–78.
[9] J.-M. Deshouillers and H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. Math. 70 (1982), 219–288.
[10] C. Hooley, *Applications of Sieve Methods*, Cambridge Tracts 70.
[11] M. N. Huxley, *Cheeger's inequality with a boundary term*, Comment. Math. Helv. 58 (1983), 347–354.

[12] M. N. Huxley, *The smallest eigenvalue corresponding to an odd function*, preprint.
[13] –, *Nodal lines and automorphic functions*, MS.
[14] H. Jeffreys and B. S. Jeffreys, *Methods of mathematical physics*, Cambridge 1962.
[15] T. Kubota, *Elementary theory of Eisenstein series*, New York 1973.
[16] N. V. Kuznetsov, *Petersson's hypothesis for forms of weight zero and Linnik's hypothesis*, Complex Science Research Institute preprint, Khabarovsk 1977, revised, Math. Sb. 111 (1980), 334–383.
[17] S. Lang, $SL_2(R)$, Addison Wesley, Reading, 1975.
[18] H. Maass, *Über eine neue Art von automorphe Funktionen und die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichungen*, Math. Ann. 121 (1949), 141–183.
[19] W. Magnus, *Non-Euclidean tesselations and their groups*, London 1974.
[20] R. A. Rankin, *Modular forms and functions*, Cambridge 1977.
[21] R. E. Showalter, *Hilbert space methods for partial differential equations*, London 1977.
[22] I. M. Vinogradov, *An introduction to number theory*, Oxford 1955.
[23] D. Zagier, *Eisenstein series and the Riemann zeta function*, in: *Automorphic forms, representation theory and arithmetic*, Bombay 1979, pp. 275–301.

Presented to the Semester
Elementary and Analytic Theory of Numbers
September 1–November 13, 1982