LECTURE NOTES: THE SPECTRAL DECOMPOSITION OF SHIFTED CONVOLUTION SUMS

GERGELY HARCOS

For the purpose of this talk I will consider a cusp form f on the upper half-plane $\mathscr{H}$ for the full modular group $\mathrm{SL}_2(\mathbb{Z})$ which is also an eigenfunction of all the Hecke operators. The Hecke eigenvalues $\lambda_f(n)$ contain important arithmetic information that can be exploited in a number of ways, most notably through the various L-functions $L(s,f)$, or $L(s,f\times\chi)$, or $L(s,f\times g)$ that they define. In these examples one can regard f as being fixed and see what happens with the L-values as one varies the point s, or the Dirichlet character χ, or another cusp form g. The square-mean of these values exhibit interesting patterns and if the family in question is sufficiently short (which is sometimes achieved by appropriate weights) then good individual bounds also follow. These bounds in turn have applications to a number of difficult questions such as the distribution of special points on spheres or arithmetic hyperbolic surfaces.

At the technical core of the study of L-values often are shifted convolution sums

$$\sum_{m-n=h}\lambda_f(m)\lambda_f(n)W(m,n)$$

where h is a positive integer and $W:(0,\infty)^2\to\mathbb{C}$ is a nice weight function. On the one hand, one expects lots of cancellation among the terms of the sum; on the other hand, one expects that the shift parameter h does not influence the size of the sum too much. In fact one aims at establishing these properties in rigorous form.

Let me start with an easier variant of the above sums. I will consider an arithmetically normalized holomorphic cusp form f of weight k,

$$f(z)=\sum_{n\geqslant 1}\lambda_f(n)n^{\frac{k-1}{2}}e(nz),$$

then f^2 is a holomorphic cusp form of weight $2k$ with Fourier decomposition

$$f^2(z)=\sum_{h\geqslant 1}\left(\sum_{m+n=h}\lambda_f(m)\lambda_f(n)(mn)^{\frac{k-1}{2}}\right)e(hz).$$

We can write $f^2(z)$ as a unique linear combination of the normalized Hecke cusp forms of weight $2k$,

$$f^2(z)=\sum_{j=1}^{\theta(k)}c_j g_{j,k}(z)$$

which yields, on the level of Fourier coefficients,

$$\sum_{m+n=h}\lambda_f(m)\lambda_f(n)(mn)^{\frac{k-1}{2}}=h^{k-\frac{1}{2}}\sum_{j=1}^{\theta(k)}c_j\lambda_{j,k}(h). \tag{1}$$

This is what I call a spectral decomposition of a convolution sum. Note that typically the terms on the left hand side are of size h^{k-1}, hence a trivial bound for the convolution sum would be $\ll_{\varepsilon,f} h^k$. However, in the light of the above decomposition and the famous bound of Deligne (1973),

$$|\lambda_{j,k}(h)| \leqslant d(h),$$

we can detect square-root cancellation among the terms:

$$\sum_{m+n=h} \lambda_f(m)\lambda_f(n)(mn)^{\frac{k-1}{2}} \ll_{\varepsilon,f} h^{k-\frac{1}{2}+\varepsilon}.$$

The adoption of the above idea to the original shifted convolution sums

$$\sum_{m-n=h} \lambda_f(m)\lambda_f(n)W(m,n)$$

raises a number of technical difficulties. First of all, we would like to consider $|f|^2$ in place of f^2 but this is no longer a holomorphic modular form. A slightly better choice is

$$F(x+iy) := y^k|f(x+iy)|^2$$

which is invariant under $\Gamma := \mathrm{SL}_2(\mathbb{Z})$ but even then we need to worry about the weight function $W(m,n)$. I will now discuss a satisfactory treatment by Good (1981) which originated from ideas by Selberg (1965). The crucial initial observation is that F lies in $L^2(\Gamma\backslash\mathscr{H})$, therefore it decomposes into a linear combination of L^2-normalized Hecke–Maass cusp forms g_j and Eisenstein series $E(\cdot,\frac{1}{2}+it)$. In particular, for the Poincaré series

$$P_h(z,s) := \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \Im(\gamma z)^s e(h\Re(\gamma z))$$

we have a Plancherel identity

$$\langle F,P_h(\cdot,s)\rangle = \sum_{j=1}^{\infty}\langle F,g_j\rangle\langle g_j,P_h(\cdot,s)\rangle + \frac{1}{4\pi}\int_{-\infty}^{\infty}\langle F,E(\cdot,\tfrac{1}{2}+it)\rangle\langle E(\cdot,\tfrac{1}{2}+it),P_h(\cdot,s)\rangle\,dt.$$

We can elaborate on the left hand side by the famous Rankin–Selberg unfolding technique:

$$\begin{aligned}
\langle F,P_h(\cdot,s)\rangle &= \int_{\Gamma\backslash\mathscr{H}} y^k|f(x+iy)|^2 \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\Im(\gamma z)^s e(-h\Re(\gamma z))\frac{dxdy}{y^2}\\
&= \int_{\Gamma_\infty\backslash\mathscr{H}} |f(x+iy)|^2 e(-hx)y^{s+k}\frac{dxdy}{y^2}\\
&= \int_0^1\int_0^\infty \sum_{m,n=1}^{\infty}\lambda_f(m)\lambda_f(n)(mn)^{\frac{k-1}{2}}e((m+n)z)e(-hx)y^{s+k-2}\,dx\,dy\\
&= \sum_{m-n=h}\lambda_f(m)\lambda_f(n)(mn)^{\frac{k-1}{2}}\int_0^\infty e^{-2\pi(m+n)y}y^{s+k-1}\frac{dy}{y}\\
&= \frac{\Gamma(s+k-1)}{(2\pi)^{s+k-1}}\sum_{m-n=h}\frac{\lambda_f(m)\lambda_f(n)}{(m+n)^s}\left(\frac{\sqrt{mn}}{m+n}\right)^{k-1}.
\end{aligned}$$

By a similar unfolding we can evaluate the terms $\langle g,P_h(\cdot,s)\rangle$ on the right hand side of the Plancherel identity in terms of gamma factors and the corresponding Hecke eigenvalues $\lambda_g(h)$, so at the end we obtain a spectral decomposition analogous to (1):

$$\sum_{m-n=h}\frac{\lambda_f(m)\lambda_f(n)}{(m+n)^s}\left(\frac{\sqrt{mn}}{m+n}\right)^{k-1} = h^{\frac{1}{2}-s}\left\{\sum_{j=1}^{\infty}\lambda_j(h)w_j(s)+\frac{1}{4\pi}\int_{-\infty}^{\infty}h^{-it}\sigma_{2it}(h)w(s,t)\,dt\right\}. \tag{2}$$

In particular, the precise form of the weight functions $w_j(s)$ and $w(s,t)$ show that the left hand side has analytic continuation in the half-place $\Re(s) > \frac{1}{2}$ and there it can be estimated as

$$\ll_\varepsilon \; h^{\frac{1}{2}-\Re(s)+\varepsilon} \sup_{j\geqslant 1} |\lambda_j(h)| \left\{ \sum_{j=1}^{\infty} |w_j(s)| + \int_{-\infty}^{\infty} |w(s,t)|\, dt \right\}, \qquad \Re(s) > \frac{1}{2} + \varepsilon.$$

The parenthesis here was estimated by Good (1981) as $\ll_{\varepsilon,f} |s|^A$ for some $A > 0$, so that by Mellin inversion the weights $(m+n)^{-s}$ on the left hand side of (2) can be replaced by any sufficiently smooth weights in m and n. The crucial ingredient here is the precise bound, established by Good (1981) for some $B > 0$,

$$\sum_{|t_j|\leqslant T} |\langle F, g_j\rangle|^2 e^{\pi |t_j|} + \int_{-T}^{T} |\langle F, E(\cdot, \tfrac{1}{2}+it)\rangle|^2 e^{\pi|t|}\, dt \ll_f T^B, \tag{3}$$

where $\frac{1}{4} + t_j^2$ (resp. $\frac{1}{4} + t^2$) is the Laplacian eigenvalue of g_j (resp. $E(\cdot, \frac{1}{2}+it)$).

Let us turn to the even more difficult case of a Hecke–Maass cusp form

$$f(x+iy) = \sqrt{y} \sum_{n\neq 0} \lambda_f(|n|) K_{ir}(2\pi|n|y) e(nx);$$

here $\frac{1}{4} + r^2$ denotes the Laplacian eigenvalue. In this case only approximate spectral decompositions were known although Sarnak (1994) and Jutila (1996) showed that (3) still holds true for the analogous function $F(x+iy) := |f(x+iy)|^2$. The failure to derive an exact formula originated from the circumstance that one only has an approximation

$$\int_0^\infty K_{ir}(2\pi my) K_{ir}(2\pi ny) y^s \frac{dy}{y} \approx \frac{G(s,ir)}{(m+n)^s} \left(\frac{\sqrt{mn}}{m+n} \right)^{2ir},$$

where $G(s,ir)$ depends only on s and ir. For sufficiently large m and n this approximation is rather precise and useful, but it can be shown that many weight functions $W(m,n)$ are missing from the linear span of the left hand side (as evaluated at various s). In fact we are dealing with infinitely many missing harmonics here.

I will now briefly report on recent progress by Blomer and myself. Using ideas from representation theory we have established an exact spectral decomposition

$$\begin{aligned} &\sum_{m-n=h} \frac{\lambda_f(m)\lambda_f(n)}{(m+n)^s} \left(\frac{\sqrt{mn}}{m+n} \right)^{100} = \\ &\quad h^{\frac{1}{2}-s} \left\{ \sum_{j=1}^{\infty} \lambda_j(h) w_j(s) + \sum_{k=1}^{\infty} \sum_{j=1}^{\theta(k)} \lambda_{j,k}(h) w_{j,k}(s) + \frac{1}{4\pi} \int_{-\infty}^{\infty} h^{-it} \sigma_{2it}(h) w(s,t)\, dt \right\}, \end{aligned} \tag{4}$$

where

$$\sum_{j=1}^{\infty} |w_j(s)| + \sum_{k=1}^{\infty} \sum_{j=1}^{\theta(k)} |w_{j,k}(s)| + \int_{-\infty}^{\infty} |w(s,t)|\, dt \ll_{\varepsilon,f} |s|^{22}, \quad \frac{1}{2} + \varepsilon < \Re(s) < \frac{3}{2} - \varepsilon. \tag{5}$$

We can see that in the decomposition not only the Maass and Eisenstein spectra participate but also the holomorphic spectrum. This explains the missing harmonics alluded to above. The theorem works for holomorphic and Maass cusp forms equally and for mixed products $\lambda_f(m)\lambda_g(n)$ as well. This harmonizes with the spectral decomposition that Motohashi (1994) proved for additive divisor sums involving $\sigma_\alpha(m)\sigma_\beta(n)$. The exponent 100 has no specific role here apart from being sufficiently large. By increasing this exponent the inequality (5) could be sharpened considerably.

The main new idea in our proof is the fact that the Hecke eigenvalues $\lambda_f(|n|)$ appear not only in the original Maass cusp form f but also in its Maass shifts. The Maass shift operators were discovered by Maass (1953) himself and its relevant properties are as follows. For every integer p there is a Maass cusp form of weight $2p$ on $\mathscr{H}$ with some Fourier decomposition

$$f_p(x+iy) = \sum_{n\neq 0} \frac{\lambda_f(|n|)}{\sqrt{|n|}} W_{f,p}(ny)e(nx).$$

Moreover, it was discovered by Kirillov (1963) that the functions $W_{f,p}$ form a complete orthogonal system in the Hilbert space $L^2(\mathbb{R}^\times)$. The functions f_p are not Γ-invariant but they can be lifted to functions in $L^2(\Gamma\backslash G)$, where $G := \mathrm{SL}_2(\mathbb{R})$:

$$\phi_p(n(x)a(y)k(\theta)) := \sum_{n\neq 0} \frac{\lambda_f(|n|)}{\sqrt{|n|}} W_{f,p}(ny)e(nx)e^{2ip\theta},$$

$$n(x) := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a(u) := \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}, \quad k(\theta) := \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

It turns out that for a suitable positive constant C_f we have a Hilbert space embedding of $L^2(\mathbb{R}^\times)$ into $L^2(\Gamma\backslash G)$ given by

$$W := \sum_{p\in\mathbb{Z}} c_p W_{f,p} \quad \longmapsto \quad \phi := C_f \sum_{p\in\mathbb{Z}} c_p \phi_p, \tag{6}$$

and the functions ϕ so obtained form an irreducible subspace of $L^2(\Gamma\backslash G)$ with the right G-action. The corresponding functions W form the Kirillov model of this automorphic representation. If $W \in L^2(\mathbb{R}^\times)$ is sufficiently smooth (which includes good behavior at 0 and $\pm\infty$) then the sums in (6) converge fast and $\phi \in L^2(\Gamma\backslash G)$ is sufficiently smooth as well. In particular, for any nice weight function $W \in L^2((0,\infty))$ we have nice vectors $\phi^\pm \in L^2(\Gamma\backslash G)$ such that

$$\phi^\pm(n(x)a(y)) = \sum_{n\geqslant 1} \frac{\lambda_f(n)}{\sqrt{n}} W(ny)e(\pm nx).$$

We decompose spectrally the product $\phi^+\phi^-$ in $L^2(\Gamma\backslash G)$ and extract an identity for the Fourier coefficients

$$\sum_{\substack{m,n\geqslant 1\\ m-n=h}} \frac{\lambda_f(m)\lambda_f(n)}{\sqrt{mn}} W(my)W(ny) =$$

$$\sum_{j=1}^\infty \frac{\lambda_j(h)}{\sqrt{h}} W_j(hy) + \sum_{k=1}^\infty \sum_{j=1}^{\theta(k)} \frac{\lambda_{j,k}(h)}{\sqrt{h}} W_{j,k}(hy) + \frac{1}{4\pi}\int_{-\infty}^\infty \frac{h^{-it}\sigma_{2it}(h)}{\sqrt{h}} W(hy,t)\,dt.$$

The weight functions on the right hand side decay very fast at infinity and they decay like the square-root function at zero which gives rise to square-root cancellation among the terms on the left hand side. Finally, we average over various weight functions W and points $y > 0$ by a technique involving the Laplace transform, and we arrive at the spectral decomposition (4) and the bound (5).

ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS, HUNGARIAN ACADEMY OF SCIENCES, POB 127, BUDAPEST H-1364, HUNGARY

E-mail address: gharcos@renyi.hu