AFL WORKING SEMINAR: OVERVIEW OF THE PROOF

WEI ZHANG,
NOTES BY TONY FENG

Contents

We're going to describe two statements that are parallel in some sense.

1. Jacquet-Rallis Fundamental Lemma

The Jacquet-Rallis fundamental lemma is a "relative" fundamental lemma, not the Langlands-Shelstad fundamental lemma proved by Ngô. (It does imply the Langlands-Shelstad Fundamental Lemma for unitary groups.)

It gives an equality of two orbital integrals: one on a unitary group, and one on GL_n. We will content ourselves with describing the objects that appear, deferring a precise formulation to a later talk.

1.1. **Local picture.** Let F/F_0 be a quadratic extension of non-archimedean local fields.[1]

There are two isomorphism classes of Hermitian spaces of given dimension for F/F_0, which we call "split" and "non-split". Let V be a Hermitian space of dimension n which is split. We consider

$$\{\text{self-dual } \Lambda \subset V\} \cong G/K$$

where $G = U(V)$ and K is a maximal compact subgroup.

We are interested in the fixed points for g – this is a version of "affine Springer fiber".

$$\mathcal{X}_g := (G/K)^g = \{\text{self-dual } \Lambda \subset V : g\Lambda = \Lambda\}.$$

Let $u \in V$. We consider imposing one more condition: $u \in \Lambda$. Then the orbital integral for g, u is

$$\mathrm{Orb}(g, u) := \#\{\Lambda \subset V \text{ self-dual } : g\Lambda = \Lambda, u \in \Lambda\}. \tag{1.1}$$

This is a finite number if (g, u) is "regular semisimple" in the invariant-theoretic sense.

Remark 1.1. In this case, the meaning of "regular semisimple" can be explicated as follows. Consider the diagonal action of G on $G \times V$ (through conjugation on the first factor). In fact it's better to linearize this: then we have a map

$$(\mathfrak{g} \times V)/G \to \mathbb{A}^?$$

Date: September 16, 2019.

[1] In the literature one sees various other notations: E/F, F'/F, ...

sending $(g,u) \mapsto (a,b)$ where $a \in F[T]^{\deg = n}$ is the characteristic polynomial of g, and $b = \langle g^i u, u\rangle$, $0 \le i \le n-1$. Concretely, regular semisimple is equivalent to $\{g^i u\}_{i=0}^{n-1}$ forming a basis for V.

Remark 1.2. The case where $u = 0$ recovers the orbital integrals in the Langlands-Shelstad fundamental lemma, for G.

The Jacquet-Rallis fundamental lemma says that

$$\mathrm{Orb}(g,u) = \text{“certain orbital integral on } \mathrm{GL}_n\text{”}.$$

We're not going to explain the actual statement. The point we want to make is that there is an *inductive* way to compute the LHS.

“Theorem”: You can compute $\{\mathrm{Orb}(g,u) \colon (g,u) \text{ r.s.}\}$ “inductively”.

Locally, there are two situations where one can perform a reduction step.

(1) (“Dimension reduction”) If u has unit norm, then the computation of $\mathrm{Orb}(g,u)$ for (g,u) is reduced to orbital integrals for unitary groups of rank $n-1$. (The point is that the lattice is a direct sum of $\langle u\rangle$ and an orthogonal complement of dimension $n-1$.)
(2) If the order $\mathcal{O}_F[g] \subset F[g]$ is actually a *maximal* order, then you can reduce to dimension $n = 1$, at the cost of replacing F/F_0 by an extension $F[g]/F_0[g]$.

These are actually both easy reductions, if you have some experience with orbital integrals.

1.2. **Global picture.** A further ingredient is needed for the induction, which is *global.* Now we let F/F_0 be a CM extension, so F_0 is totally real. Pretend for simplicity that it's unramified everywhere. Consider

$$\mathsf{Lat}_n := \{\Lambda = \text{self-dual Hermitian rank } n \text{ lattice } \mid \text{positive definite}\}/\sim$$

This is the analogue of the affine Grassmannian.

Let's reformulate the local picture a bit. The local counting was taking place within G/K, the “affine Grassmannian”. The “affine Springer fiber” is $\mathcal{X}_g := (G/K)^g$. We could then define a “special divisor” $Z(u) := \{\Lambda \colon u \in \Lambda\} \subset G/K$. Then the orbital integral (1.1) can be interpreted as

$$\mathrm{Orb}(g,u) = \#(G/K)^g \cap Z(u).$$

Globally, the analogue is as follows. Fix $a \in F[T]_{\deg n}$. (We'll usually take a to be irreducible.)

- We define
$$\mathrm{CM}(a) := \{(\Lambda, \varphi) \colon \operatorname{char}(\varphi) = a\}$$
which is analogous to $\mathcal{X}_g$. This is a locus of lattices with extra endomorphisms.
- We define
$$Z(m) = \left\{ (\Lambda, u) \colon \begin{matrix} u \in \Lambda \\ \langle u, u\rangle = m \end{matrix} \right\}.$$
Note that now the u is varying.

Globally we are interested in

$$\mathrm{CM}(a) \cap Z(m) = \left\{ (\Lambda \in \mathsf{Lat}_n, \varphi \in \mathrm{End}(\Lambda), u \in \Lambda) \colon \begin{matrix} \operatorname{char}(\varphi) = a \\ \langle u, u\rangle = m \end{matrix} \right\}.$$

A global-local argument shows that $\#\,\mathrm{CM}(a) \cap Z(m)$ is related to $\prod_v \mathrm{Orb}(g,u)$. More precisely, it can be equated with an expression of the form

$$\sum_{(g,u)} \mathrm{Orb}(g,u) \tag{1.1}$$

where the sum runs over (g,u) having invariant (a,b) such that $b_0 = m$ (note that the orbital integral really only depends on the (a,b)). The point is that the local counting problem is embedded in the global counting problem. In fact by formulating a more general statement with ramification allowed, we can isolate a single term, as is familiar in the study of the trace formula.

Question: can you globalize a non-maximal local order to a global order that is maximal at *all other* places? Morally we'd like to do this, to control all the orbital integrals except the factor that comes from p in (1.1). But we will do it in a way that doesn't require us to answer this question, which seems hard. This allows us to pretend we understand what goes on away from p.

Now, how do you go from the case where m is a p-adic unit to everything? We use a "trivial form of Ihara's Lemma". There are two ways of embedding $M_k(\Gamma) \to M_k(\Gamma_0(p))$. Hence we get

$$M_k(\Gamma)^{\oplus 2} \to M_k(\Gamma_0(p))$$

and "Ihara's Lemma over $\mathbf{C}$" says it's injective. This is easy; when people say "Ihara's lemma" they are usually referring to the much subtler case where you take $\mathbf{F}_p$-coefficients. Anyway, this allows you to prove:

Corollary 1.3. *Let $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ (or more generally, a congruence subgroup which has no level at p). Suppose*

$$f = \sum_{n \geq 0} a_n q^n \in M_k(\Gamma),$$

with $k > 0$, such that $a_n = 0$ for all $(n,p) = 1$. Then $f = 0$.

Theorem 1.4. *The generating series*

$$\sum_m \#\,\mathrm{CM}(a) \cap Z(m) q^m$$

is modular, i.e. lies in $M_k(\Gamma)$.

In fact,

$$\sum_m (\#\,\mathrm{CM}(a) \cap Z(m)) q^m = \sum_{(\Lambda,\varphi) \in \mathrm{CM}(a)} \theta_\Lambda$$

where

$$\theta_\Lambda = \sum_{u \in \Lambda} q^{\langle u,u \rangle}$$

is the θ function attached to Λ. Corollary 1.3 shows that this modular form is determined by the u such that $\langle u,u \rangle$ is a unit at p.

2. The Arithmetic Fundamental Lemma

2.1. **Local picture.** The formulation of the AFL is analogous.

- The affine Grassmannian gets replaced by *Rapoport-Zink space* $\mathcal{N}_n$.
- $Z(u)$ gets replaced by the *Kudla-Rapoport divisor* $Z(u) \hookrightarrow \mathcal{N}_n$, and
- The affine Springer fiber gets replaced by a cycle $\mathcal{N}_n^g \hookrightarrow \mathcal{N}_n$.

The AFL is about $\mathcal{N}_n^g \cap Z(u)$:

$$\mathrm{Int}(g,u) := \chi(\mathcal{N}_n, \mathcal{O}_{\mathcal{N}_n^g} \otimes^{\mathbb{L}} \mathcal{O}_{Z(u)}) \in \mathbf{Z} \tag{2.1}$$

for (g,u) regular semisimple.

The two local recursive properties still hold, but are more difficult to establish.

2.2. **Global picture.** The global definition is actually easier. The role of Lat_n is played by $\mathcal{M}_n$, an integral model of a certain PEL Shimura variety. Roughly,

$$\mathcal{M}_n = \left\{ (A,\iota,\lambda) \colon \begin{matrix} A = \text{abelian variety} \\ \iota \colon \mathcal{O}_F \to \mathrm{End}(A), \text{ sign} = (n-1,1) \\ \lambda \colon A \xrightarrow{\sim} A^\vee \end{matrix} \right\}$$

(In fact Lat_n could be viewed as the "positive definite" case, where the signature is instead $(n,0)$.) The map $\mathcal{M}_n \to \operatorname{Spec} \mathcal{O}_F$ has relative dimension $n-1$.

Again we can define $\mathrm{CM}(a)$ and $Z(m)$. The $Z(m)$ is the global "Kudla-Rapoport divisor". One again forms the generating series

$$\sum_{m \geq 0} Z(m) q^m \in \mathrm{CH}^1(\mathcal{M}_n)_{\mathbf{Q}} \otimes \mathbf{Q}[[q]]$$

and a Theorem [BHKRY] asserts that it is a modular form:

$$\sum_{m \geq 0} Z(m) q^m \in \mathrm{CH}^1(\mathcal{M}_n)_{\mathbf{Q}} \otimes \in M_n(\Gamma). \tag{2.1}$$

In fact one needs to promote this to the "arithmetic Chow group" $\widehat{\mathrm{CH}}^1(\mathcal{M}_n)_{\mathbf{Q}} \xrightarrow{\sim} \widehat{\mathrm{Pic}}^1(\mathcal{M}_n)$.

This is really a generalization of the fact that theta functions are modular, e.g.

$$\sum_m \#(\Lambda_0 \cap Z(m)) q^m = \theta_{\Lambda_0}$$

where $\Lambda_0 \cap Z(m)$ is the subset of $\Lambda \in Z(m)$ where $\Lambda = \Lambda_0$.

The definition of $\mathrm{CM}(a)$ is pulled back from the Siegel case. Over $\mathcal{A}_g$, we define

$$\mathrm{CM}(a) := \{(A, \lambda, \varphi \in \mathrm{End}^0(A)) \colon \operatorname{char}(\varphi) = a\}$$

The characteristic polynomial $\operatorname{char}(\varphi)$ has degree $2g$ here; it can be defined as the characteristic polynomial of φ on the ℓ-adic Tate module of A, which will in fact have coefficients in $\mathbf{Q}$. If $\mathcal{O}_F[a] \subset F[a]$ is a maximal order, then $\mathrm{CM}(a)$ will be "horizontal" curve, flat over $\operatorname{Spec} \mathbf{Z}$. If a is irreducible, it always generically has dimension 0 but if $\mathcal{O}_F[a]$ is not maximal, then some fibers will be positive-dimensional (hence "big fat CM cycle").

We replace $\mathrm{CM}(a)$ with a derived version ${}^{\mathbb{L}}\mathrm{CM}(a)$. The arithmetic intersection is a pairing

$$\widehat{\mathrm{CH}}^1(\mathcal{M}_n) \times Z_1(\mathcal{M}_n) \to \mathbb{R}. \tag{2.2}$$

The pairing works by interpreting $\widehat{\mathrm{CH}}^1(\mathcal{M}_n)$ as $\widehat{\mathrm{Pic}}^1(\mathcal{M}_n)$, the group of arithmetic line bundles; one can then restrict an arithmetic line bundle to a curve and form its degree. The pairing doesn't factor through $\mathrm{CH}_1(\mathcal{M}_n)$. However it does factor through the quotient by rationally trivial cycles supported on special fibers.

One then has to argue that ${}^{\mathbb{L}}\mathrm{CM}(a) \in Z_1(\mathcal{M}_n)$. Then it makes sense to pair it with (2.1), and it is related to the local intersection numbers (2.1) $\mathrm{Int}(g,u)$ at $p < \infty$. You also have to analyze archimedean local heights. Then you use modularity in a similar way.

Remark 2.1. There's a special feature of the arithmetic situation which makes things actually easier. We can regard the pairing (2.2) as a map of $\mathbf{Q}$-vector spaces, so in particular you view $\mathbb{R}$ as a $\mathbf{Q}$-vector space. The pairing lands in $\sum_p \mathbf{Q} \log p$ plus an archimedean thing. If you can compute the archimedean thing, the independence of $\log p \in \mathbb{R}$ allows you to *automatically* separate contributions from different primes.

The last step is to prove a local constancy of intersection numbers. A simplification for this step was given by Mihatsch.