Gross–Zagier formula and arithmetic fundamental lemma

Wei Zhang

Abstract. The recent work [**YZZ**] completes a general Gross–Zagier formula for Shimura curves. Meanwhile an arithmetic version of Gan–Gross–Prasad conjecture proposes a vast generalization to higher dimensional Shimura varieties. We will discuss the arithmetic fundamental lemma arising from the author's approach using relative trace formulae to this conjecture.

Contents

1. Introduction

In this survey article we discuss some global restriction problems for automorphic representations proposed by Gan–Gross–Prasad and their arithmetic version. The global restriction problem is a generalization of a result of Waldspurger (among other works). The arithmetic version is on certain algebraic cycles on unitary Shimura varieties and it is a natural generalization of the Gross–Zagier formula for modular and Shimura curves. It will have numerous arithmetic applications such as to the Beilinson–Bloch conjecture which is a generalization of the Birch–Swinnerton-Dyer conjecture. For the restriction problem, Jacquet–Rallis proposed an approach using relative trace formula to attack the unitary case of Gan–Gross–Prasad conjecture. Later on the author proposed a strategy to attack the arithmetic version for unitary Shimura varieties. In this article, we will only discuss the first step of this approach, namely, an *arithmetic fundamental lemma* which is an equality between certain intersection numbers and the first derivatives of some relative orbital integrals. The fundamental lemma arising from Jacquet–Rallis's approach

1991 *Mathematics Subject Classification.* Primary 11G18, 11G50; Secondary 22E55.
Key words and phrases. Gross–Zagier formula, Gan–Gross–Prasad conjecture, relative trace formula, arithmetic fundamental lemma.
The author is supported in part by NSF Grant DMS #1001631.

has been settled by Yun. The arithmetic fundamental lemma however remains open except in some lower dimensional cases which were verified by the author by explicit computations.

We outline the contents of the article. We will first describe the global restriction problems. Then we recall the joint work with X. Yuan and S. Zhang on Gross–Zagier formula ([**YZZ**]) which serves as a prototype of the more general arithmetic restriction problems (stated in e.g., [**Zh4**]). We then present the relative trace formula approach to the restriction problems. In the last section we state the intersection problem on unitary Rapoport–Zink spaces and the arithmetic fundamental lemma.

Some important issues are untouched in this survey article. For example, we choose only to present results of equivalence of non-vanishing of periods/heights and L-values. An equality between them will be also crucial for further number theoretical investigation. In the case of periods, Waldspurger's formula ([**Wa**]) is a prototype and Ichino–Ikeda ([**II**]) proposed a refined version of the global conjecture of Gan–Gross–Prasad ([**GGP**]). Some other important work related to the questions in this survey include [**I**],[**GJR1**],[**GJR2**] etc..

Acknowledgement *The author would like to thank Professor Shouwu Zhang for some comments on an early version. The author thanks the Morningside Center of Mathematics and Mathematical Sciences Center of Tsinghua University for their hospitality where some part of the work was done.*

2. Restriction problem: period and height

2.1. Automorphic period. We may start with a classical question. Let F be a number filed and let $\mathbb{A}$ be the ring of adeles. Let G be a reductive group over F, and let $H \subset G$ be a (reductive) subgroup. Let Z_G be the center of G and let $Z = H \cap Z_G$. Let $\mathcal{A}_0(G)$ be the space of cuspidal automorphic forms on $G(\mathbb{A})$. Then the *automorphic H-period* is defined by

$$\ell_H : \mathcal{A}_0(G) \to \mathbb{C}$$
$$\phi \mapsto \int_{[H]} \phi(h)dh.$$

where

$$[H] := Z(\mathbb{A})H(F)\backslash H(\mathbb{A}).$$

Let π be a cuspidal automorphic representation with an embedding $\pi \subset \mathcal{A}_0(G)$. For simplicity, we assume the multiplicity $m(\pi)$ is one so that the embedding is unique. We consider the restriction OF ℓ_H to π and the question we may ask is when it does not vanish:

$$\ell_{H,\pi} \neq 0\,?$$

One obvious obstruction for the nonvanishing of H-period is the nonvanishing of the space of H-invariant linear functionals

$$\ell_{H,\pi} \in \mathrm{Hom}_{H(\mathbb{A})}(\pi, \mathbb{C}) \neq 0.$$

Sometimes it is also necessary to consider a twisted version. Namely, we insert a character of $H(\mathbb{A})$ in the definition of H-period and we call the new integral $\ell_{H,\chi}$ the (H, χ)-period.

The subgroup H need not to be reductive but we will assume so since only reductive subgroups will appear in all examples of this article.

Example 2.1. This was a special case studied by Waldspurger ([**?**]) and revisited by Jacquet ([**J**]). Let E/F be a quadratic extension of number fields. Let B be a quaternion algebra over F together with an embedding $E \subset B$. We then the induced embedding of torus $T = E^\times$ into $G = B^\times$. For a cuspidal automorphic representation π of $G(\mathbb{A})$, the restriction problem is to determine when the toric period $\ell_{T,\pi}$ does not vanish.

Example 2.2. The major example we will discuss is one of the restriction problems of Gan–Gross–Prasad. Let E/F be a quadratic extension of number fields. Let V be a(non-degenerate) Hermitian space and let $W \subset V$ be a subspace of codimension one. Denote by $U(V), U(W)$ the unitary groups respectively with an embedding $U(W) \subset U(V)$. Then the restriction problem concerns the pair $G = U(W) \times U(V)$ and its subgroup H of the diagonal embedding of $U(W)$. Gan–Gross–Prasad made a conjecture on when the period is non-zero in terms of the central value of a suitable Rankin–Selberg L-function. Later we will recall Jacquet–Rallis's approach to this conjecture using relative trace formulae.

2.2. Heights of cycles on Shimura varieties. We now discuss the arithmetic version of the global restriction problems. Suppose that both the reductive group G and its subgroup H can define Shimura varieties and the Shimura data are compatible so we have an embedding

$$\mathrm{Sh_H} \longrightarrow \mathrm{Sh_G}.$$

The Shimura variety $\mathrm{Sh_G}$ is a projective system $\mathrm{Sh_{G,K}}$ indexed by open compact $K \subset G(\mathbb{A}^\infty)$. For simplicity we assume both are projective over a number field. Otherwise we may take some smooth compactification. On the Shimura variety $\mathrm{Sh_G}$ there is an action of $G(\mathbb{A}^\infty)$ where $\mathbb{A}^\infty$ denotes the finite adeles. We then have a cycle class map from the Chow group (with rational coefficient) to the cohomology group

$$\mathrm{cl}_* : \mathrm{Ch}^*(\mathrm{Sh_G}) \to \mathrm{H}^{2*}(\mathrm{Sh}_G)$$

compatible with the $G(\mathbb{A}^\infty)$-action. Here we may take as $\mathrm{H}^{2*}(\mathrm{Sh}_G)$ any Weil cohomology with suitable coefficients. And the Chow group and cohomological group are inductive limits of that of $\mathrm{Sh_{G,K}}$ indexed by open compact $K \subset G(\mathbb{A}^\infty)$. we then have a two-step filtration on the Chow group

$$(\star) \qquad 0 \subset \mathrm{Ch}^*(\mathrm{Sh_G})_0 \subset \mathrm{Ch}^*(\mathrm{Sh_G})$$

where $\mathrm{Ch}^*(\mathrm{Sh_G})_0$ is the kernel of the cycle class map. For a $G(\mathbb{A}^\infty)$-module π, we are interested in the π-isotypical component

$$\mathrm{Ch}^*(\mathrm{Sh_G})_0[\pi] := \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi, \mathrm{Ch}^*(\mathrm{Sh_G})_0) \otimes \pi$$

which will be naturally considered as a subspace of $\mathrm{Ch}^*(\mathrm{Sh_G})_0$.

According to a series of conjectures of Beilinson and Bloch on the filtration of Chow groups of a smooth projective variety X defined over a filed k, $\mathrm{Ch}^i(X)_\mathbb{Q}$ has a filtration of $i+1$ steps

- The j-th graded piece is "controlled" by $H^{2i-j}(X)$.
- the 0-th graded piece of this filtration is the image of the cycle class map which is controlled by $H^{2i}(X)$ by the Hodge-type (or Tate-type) conjecture for the corresponding Weil cohomology.

Moreover, they conjecture that this filtration has only two steps if k is a number field, namely given by ($\star$) defined above! Assuming the conjecture of Beilinson and Bloch, since our Shimura varieties are certainly defined over number fields, the $\mathrm{Ch}^*(\mathrm{Sh_G})_0$ should be controlled by $\mathrm{H}^{2*-1}(\mathrm{Sh}_G)$. This has the following consequence:

For a $G(\mathbb{A}^\infty)$-module π, the π-isotypical component $\mathrm{Ch}^*(\mathrm{Sh_G})_0[\pi]$ is zero unless $\mathrm{H}^{2*-1}(\mathrm{Sh_G})[\pi] \neq 0$.

Now we assume further that

- The Shimura subvariety $\mathrm{Sh_H}$ is in the arithmetical middle dimension of $\mathrm{Sh_G}$, i.e.:
$$\dim \mathrm{Sh_H} = \frac{1}{2}(\dim \mathrm{Sh_G} - 1).$$
- The $G(\mathbb{A}^\infty)$-equivariant exact sequence
$$0 \longrightarrow \mathrm{Ch}^*(\mathrm{Sh_G})_0 \longrightarrow \mathrm{Ch}^*(\mathrm{Sh_G}) \longrightarrow \mathrm{Im}(\mathrm{cl}_*) \longrightarrow 0$$
split (after extending coefficients to a large field, say $\mathbb{C}$) when $*$ is the codimension of $\mathrm{Sh_H}$.

Similarly we may consider $\mathrm{Ch}_*(\mathrm{Sh_G})$, the *projective* limit of the homological Chow group (for more details see [**Zh4**]). And $\mathrm{Sh_H}$ defines a class in $\mathrm{Ch}_*(\mathrm{Sh_G})$ for $* =$ dim $\mathrm{Ch}_*(\mathrm{Sh_G})$. We denote by $[\mathrm{Sh_H}] \in \mathrm{Ch}_*(\mathrm{Sh_G})_0$ the projection of the class of $\mathrm{Sh_H}$. We do not need the full strength of the second part; all we need is the cohomological trivialization $[\mathrm{Sh_H}]$ of $\mathrm{Sh_H}$.

Then using the (conditionally defined) Beilinson–Bloch height pairing, we may define a linear functional which by abuse of notation is still denoted by ℓ_H
$$\begin{aligned}\ell_H : \mathrm{Ch}^*(\mathrm{Sh_G})_0 &\to \mathbb{C}\\ z &\mapsto \langle z, [\mathrm{Sh_H}]\rangle.\end{aligned}$$
This is the arithmetic version of the automorphic period.

Then the arithmetic version of the restriction problems asks: for π appearing in the cohomology $\mathrm{H}^{2*-1}(\mathrm{Sh_G})$, when is the restriction nonvanishing
$$\ell_H|_{\mathrm{Ch}^*(\mathrm{Sh_G})_0[\pi]} \neq 0 \quad ?$$
According to Beilinson–Bloch's generalization of Birch–Swinnerton-Dyer conjecture, the rank of $\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi, \mathrm{Ch}^*(\mathrm{Sh_G})_0)$ should be the vanishing order of a suitable L-function attached to π at the center of its functional equation. In our case, it is expected (at least for the two examples below) that the non-vanishing of $\ell_H|_{\mathrm{Ch}^*(\mathrm{Sh_G})_0[\pi]}$ is equivalent to the nonvanishing of the first derivative of the L-function and the nonvanishing of the space $\mathrm{Hom}_{H(\mathbb{A}^\infty)}(\pi, \mathbb{C})$.

Corresponding to the two examples for periods, we have two examples for heights.

Example 2.3. Let E be a quadratic CM extension of a totally real field F. Let B be a quaternion algebra over F together with an embedding $E \subset B$. Assume that B_∞ is indefinite at one place and definite at all other places. We then the induced embedding of torus $T = E^\times$ into $G = B^\times$ which will induces an embedding of a zero-dimensional Shimura variety $\mathrm{Sh_T}$ into a Shimura curve $\mathrm{Sh_G}$ with images called CM points. In the special case $F = \mathbb{Q}$ and $B = M_{2,\mathbb{Q}}$ is the matrix algebra (and under the Heegner conditions), this was studied by Gross–Zagier. We will discuss this example in more details in the next section.

Example 2.4. This is the arithmetic version of the global Gan–Gross–Prasad conjecture for unitary groups. Let E/F be the same as in the previous example. Let V be a(non-degenerate) Hermitian space of dimension n such that V_∞ has signature $(n-1,1)$ at one place and positive definite at all others. Let $v \in V$ be a vector whose norm is totally positive and let W be the orthogonal complement of v. Then W_∞ has signature $(n-2,1)$ at one place and positive definite at all others. We still denote the pair $G = U(W) \times U(V)$ and its subgroup H of the diagonal embedding of $U(W)$. Then we have an arithmetic version of the Gan–Gross–Prasad conjecture in the unitary case. An approach to this conjecture was first proposed by the author. In this survey, we will only discuss the first step of this approach, namely, the arithmetic fundamental lemma.

3. Gross-Zagier formula for Shimura curves

In this section we recall the joint work with Xinyi Yuan and Shouwu Zhang ([**YZZ**]) on Gross–Zagier formula. In 1984, Gross and Zagier [**GZ**] proved a formula that relates the Néron–Tate heights of Heegner points to the central derivatives of some Rankin L-series under certain ramification conditions. Since then some generalizations are given in various papers [**Zh1, Zh2, Zh3**]. The methods of proofs of the Gross–Zagier theorem and all its extensions depend on some newform theories. There are essential difficulties to remove all ramification assumptions by these methods. In [**YZZ**] by refining the methods of previous works, we remove all ramification condition to arrive at a complete Gross–Zagier formula for Shimura curves. Such a general formula is an arithmetic analogue of the central value formula of Waldspurger [**Wa**] and has been speculated by Gross [**Gr**] in 2002 in term of representation theory.

Let σ be a cuspidal automorphic representation of $GL_2(\mathbb{A})$ with finite order central character ω. Let $T = E^\times$ as an algebraic group over F. Let χ a Hecke character of $T(\mathbb{A}) = \mathbb{A}_E^\times$. We assume that

$$\chi|_{\mathbb{A}^\times} \cdot \omega = 1.$$

Denote by $L(s,\sigma,\chi)$ the Rankin-Selberg L-function. More precisely it is the L-function of the base change of σ to E then twisted by χ. It has a functional equation

$$L(s,\pi,\sigma) = \epsilon(s,\sigma,\chi)L(1-s,\sigma,\chi).$$

At the center, one has a decomposition

$$\epsilon(\frac{1}{2},\sigma,\chi) = \prod_v \epsilon(\sigma_v,\chi_v).$$

Denote

$$\Sigma = \{\text{ place } v \text{ of } F: \ \epsilon(\sigma_v,\chi_v) \neq \chi_v(-1)\eta_v(-1)\},$$

where $\eta : F^\times\backslash\mathbb{A}^\times \to \{\pm 1\}$ is the quadratic character associated to the extension E/F by class field theory. Then Σ is a finite set and the global root number is given by

$$\epsilon\left(\frac{1}{2},\sigma,\chi\right) = \prod_v \epsilon(\sigma_v,\chi_v) = (-1)^{\#\Sigma}.$$

We define for each v a quaternion algebra B_v over F_v which is division if and only if $v \in \Sigma$. Then by Theorem of Saito–Tunnell ([**Sa**],[**Tu**]), we can embed $E_v \hookrightarrow B_v$ and σ_v has a non-zero Jacquet–Langlands correspondence π_v on $B_v^\times$ such that

$$\dim \mathrm{Hom}_{T_v}(\pi_v, \chi_v) = 1.$$

The collection $\{B_v\}_v$ of algebras form a quaternion algebra $\mathbb{B}$ over $\mathbb{A}$, i.e., $\mathbb{B}$ is the unique $\mathbb{A}$-algebra which is free of rank 4 as an $\mathbb{A}$-module and whose localization $\mathbb{B}_v := \mathbb{B} \otimes_{\mathbb{A}} F_v$ is isomorphic to B_v at any place v. Such an algebra exists uniquely up to isomorphism. In fact, it is a restricted product of $\{B_v\}_v$ in the sense that

$$\mathbb{B} = \prod_v{}' B_v = \prod_{v \in \Sigma} B_v \times M_2(\mathbb{A}^\Sigma).$$

We also form a representation $\pi = \prod_v \pi_v$ of $\mathbb{B}^\times$. *Note that by choosing this π, we have killed all local obstructions, namely* $\mathrm{Hom}_{T(\mathbb{A})}(\pi, \chi) \neq 0$.

Case one: $\#\Sigma$ even. Then there exists a unique up to isomorphism quaternion algebra B over F such that $B \otimes \mathbb{A} \simeq \mathbb{B}$. Then π is in fact a cuspidal automorphic representation. Then a theorem of Waldspurger asserts that the following are equivalent

(i) The (T, χ)-period of π does not vanish: $\ell_{T,\chi,\pi} \neq 0$.
(ii) The first central value does not vanish $L(1/2, \sigma, \chi) \neq 0$.

REMARK 3.1. As mentioned earlier, Waldspurger actually proved an exact formula relating the period and the L-value.

Case two: $\#\Sigma$ odd. We need some assumptions:

- F be a totally real field and E a CM extension of F.
- For all $v|\infty$, π_v is discrete series of weight two. In particular, Σ contains all archimedean places and the Jacquet–Langlands correspondence π_v for $v|\infty$ is the trivial representation.

Then the adelic quaternion algebra $\mathbb{B}$ does not arise from the base change of any quaternion algebra over the number field F. Following Kudla's terminology we call such $\mathbb{B}$ an *incoherent* quaternion algebra.

In this case, we have a Shimura curve $\mathrm{Sh}_{\mathbb{B}}$ defined over the totally real field F. For the construction we refer to [**YZZ**]. And Sh_{T} is a zero dimensional Shimura variety which is embedded into the Shimura curve $\mathrm{Sh}_{\mathbb{B}}$. In this case, all assumptions in the previous section become unconditional so we have a well-defined linear functional $\ell_{H,\chi}$ of $\mathrm{Ch}^1(\mathrm{Sh}_{\mathbb{B}})_0$ for a finite order character χ of $T(\mathbb{A})$. Here the Beilinson–Bloch height is reduced to the Néron–Tate height pairing.

THEOREM 3.2. *The following are equivalent:*

- *The Néron-Tate height does not vanish: $\ell_{T,\chi,\pi} \neq 0$.*
- *The first central derivative does not vanish: $L'(1/2, \sigma, \chi) \neq 0$.*

Just like the Waldspurger's case, we also have an exact formula. See [**YZZ**].

REMARK 3.3. Under various conditions, this was proved earlier by Gross–Zagier and S. Zhang.

REMARK 3.4. The assumption on weights of π_∞ are not essential. But the assumption that all Σ containing all archimedean places is essential. It is a folklore open problem to look for a Gross–Zagier type formula for Maass forms. Our method does not offer any insight to this problem.

4. Relative trace formula

4.1. Relative trace formula (RTF). To study automorphic periods, one natural tool is the relative trace formula. Let G be a reductive group with two subgroups H_1, H_2. For $f \in \mathcal{C}_c^\infty(G(\mathbb{A}))$, we may consider the operator $R(f)$ on the Hilbert space $L^2(G(F)\backslash G(\mathbb{A}))$ (or with a fixed central character). It is an integral operator represented by the kernel function

$$K_f(x, y) := \sum_{\delta \in G(F)} f(x^{-1}\delta y)$$

which is $G(F)$-invariant for both $x, y \in G(\mathbb{A})$. We then define a distribution on $G(\mathbb{A})$ by

$$I(f) := \int_{[H_1]} \int_{[H_2]} K_f(h_1, h_2) dh_1 dh_2.$$

For a cuspidal automorphic representation π, we let

$$I_\pi(f) := \sum_{\phi \in \mathcal{B}(\pi)} \ell_{H_1}(\pi(f)\phi)\overline{\ell_{H_2}(\phi)}$$

for an orthonormal basis $\mathcal{B}(\pi)$ of π. Then we expect to have a relative trace formula identity of the following form

$$\sum_{\delta \in H(F)\backslash G(F)_{rs}/H(F)} \tau(\delta) O(\delta, f) + ... = \sum_\pi I_\pi(f) + ...$$

Here the RHS is the spectral side which sums over all cuspidal automorphic representations and the omitted part is the spectrum contribution from others. In the LHS, "rs" indicates "regular semisimple" in a relative sense which will be defined below. And $\tau(\gamma)$ will be a certain volume factor and $O(\gamma, f)$ is a relative orbital integral. The identity can be made rigorous if we use some simplified version of relative trace formula, at least in the example below.

Therefore, *to produce such a trace formula identity, we only need to input a triple* (G, H_1, H_2) *as above.* Usually it will also be useful if we allow to insert a character of $H_1(\mathbb{A}) \times H_2(\mathbb{A})$, as we will see in our example of Jacquet–Rallis's RTFs.

Example 4.1 (Arthur–Selberg trace formula). Let $G = H \times H$ and let $H_1 = H_2 = \Delta(H)$ be the diagonal embedding of H. Then the relative trace formula associated to this triple is essentially reduced to the Arthur–Selberg trace formula associated to H.

Our central interest is to relate some period to a suitable L-value. For this, we first hope to find another period on a different group which will essentially give the L-value. Then the question is reduced to relating two periods on two different groups. This can usually be achieved by comparing two relative trace formulae. In general, it is still quite mysterious when two RTFs are comparable. The forthcoming works of Sakellaridis ([**S**]) and Sakellaridis–Venkatesh ([**SV**]), among other things, aim to provide criterions when we can compare two RTFs.

Example 4.2. In the theorem of Waldspurger, we may compare the following two RTFs. Let E/F be a quadratic extension.

(i) (GL_2, A, A) and insert a quadratic character in the first A. Here A is the subgroup of diagonal matrices.

(ii) $(B^\times, E^\times, E^\times)$. Here B is a quaternion algebra and E can be embedded into B.

EXAMPLE 4.3. Now we present the construction of relative trace formulae by Jacquet-Rallis to attack the Gan–Gross–Prasad conjecture for unitary group.

(i) Unitary side: for a pair $W \subset V$, let $G = U(V) \times U(W)$ and let $H_1 = H_2$ be the diagonal embedding of $U(W)$ into G.
(ii) Linear side:
 – $G' = Res_{E/F}\,(GL_{n-1} \times GL_n)$
 – $H_1 = Res_{E/F} GL_{n-1}$, $H_2 = GL_{n-1,F} \times GL_{n,F}$.

 Moreover we need to insert a quadratic character of $H_1(\mathbb{A}) \to \{\pm 1\}$ given by $(h_{n-1}, h_n) \mapsto \eta(det(h_{n-1}))$ $(\eta(det(h_n))$, resp.) if n is odd (even, resp.). The quadratic character appears due to the period characterization of the image of quadratic base change from unitary group.

4.1.1. *Invariant theory.* We recall some notions from [**AG**]. Let F be a local field of characteristic zero. Let H be reductive group and V an affine F-variety together with an action of H. The categorical quotient $(V//H, \pi)$ is the affine variety $V//H := \mathrm{Spec}(\mathcal{O}(V)^H)$ together with the natural (surjective) morphism $\pi : V \to \mathrm{Spec}(\mathcal{O}(V)^H)$. *However, in the level of F-rational points $\pi(F) : V(F) \to (V//H)(F)$ need not to be surjective.*

DEFINITION 4.4. Let $x \in V(F)$. We say that x is

- *H-semisimple* (or *semisimple relative to H*) if Hx is Zariski closed in V (or equivalently, $H(F)x$ is closed in $V(F)$ for the analytic topology induced from F.)
- *H-regular* if the stabilizer of x has minimal dimension.

EXAMPLE 4.5. The notions coincide with the usual ones when we consider the adjoint representation of a reductive H on its Lie algebra $V = \mathfrak{h}$.

A stable regular semisimple orbit is by definition the preimage of a point in the categorical quotient. Any two points within one stable regular semisimple orbit are $H(\bar{F})$-equivalent, but may not $H(F)$-equivalent. In any way, a stable regular semisimple orbit contains at most finitely many $H(F)$-orbits.

In a RTF associated to a (reductive) triple (G, H_1, H_2) as above, we consider the action of $H := H_1 \times H_2$ on G given by $(h_1, h_2) \circ g = h_1^{-1} g h_2$. Then we may have notions of "semisimple" and "regular" (H is omitted if no confusion arises).

To compare two RTFs, we usually would like to match their (at least, stable regular semisimple) orbits. In practice, this can be achieved by identifying the categorical quotients. In each of the two examples below, the categorical quotients of the two RTFs involved are naturally isomorphic. A stable regular semisimple orbit has trivial stabilizer, in particular, consists of a single $H(F)$-orbit.

EXAMPLE 4.6. We first describe the categorical quotient appeared in the RTF associated to $(B^\times, E^\times, E^\times)$. There is a unique decomposition $B = B_+ + B_-$ where B_+ is the image of the embedded E and B_- is the "orthogonal complement" of B_+ for the quadratic form given by the reduced norm N on B. Then we may define a morphism to the affine line $\mathbb{A}^1_F$ over F:

$$\pi_{B,E} : B^\times \to \mathbb{A}^1$$
$$b \mapsto \frac{\mathrm{N}b_+}{\mathrm{N}b}.$$

It is a surjective morphism between varieties. Then the pair $(\pi_{B,E}, \mathbb{A}^1)$ is a categorical quotient for $B^\times//(E^\times \times E^\times)$. This construction also works for the case $E = F \times F$ in which case $B = M_{2,F}$ is unique. We then obtain the categorical quoteint for $GL_2//(A \times A)$. In both cases, a point is regular semisimple if and only if its image is not $0, 1$ in the affine line $\mathbb{A}^1$. In the level of F-rational points, $\pi_{B,E}(F) : B^\times \to \mathbb{A}^1(F) = F$ is not surjective if E is a genuine quadratic field extension of F and surjective if $E = F \times F$. However, if we take all (isomorphism classes of) quaternion algebras B which E is embedded into, then we have a disjoint union

$$\coprod_B \pi_{B,E}(B^\times_{rs}) = \mathbb{A}^1(F) - \{0,1\} = F - \{0,1\} = \mathbb{P}^1(F) - \{0,1,\infty\}$$

where the RHS is precisely the image of $\pi_{M_{2,F},F\times F}(F)$ of regular semisimple elements. Then for a $\gamma \in GL_2(F)$ regular semisimple relative to $A \times A$, and $\delta \in B^\times$ regular semisimple relative to $E^\times \times E^\times$, we say they match each other and we write $\gamma \leftrightarrow \delta$ if the images of γ, δ coincide under the respective quotient maps. This only depends on the orbit of respective subgroups and defines a bijection

$$A(F)\backslash GL_2(F)_{rs}/A(F) = \coprod_B E^\times\backslash B^\times_{rs}/E^\times.$$

Example 4.7. We now discuss the two RTFs of Jacquet–Rallis ([**Z**]). In the unitary side, assume that $W \subset V$ is the orthogonal complement of a vector $v \in V$ with norm $(v,v) = 1$. We immediately see that $G//(H \times G)$ is isomorphic to $U(V)//U(W)$ where $U(W)$ acts by conjugation. Then ring of $U(W)$-invariant regular functions on $U(V)$ is generated by: for $g \in U(V)$,

$$tr \wedge^i g, (g^j v, v), \quad i = 1,2,...,\dim V, j = 0,1,...,\dim V - 1.$$

And g is regular semisimple if and only if $det(g^i v, g^j v)_{0\le i,j\le \dim V-1} \neq 0$ (or equivalently $g^j v$ for $0 \le j \le \dim V - 1$ is a basis of V). On the linear side, we define an symmetric space over F

$$S_n = \{s \in GL_{n,E} : s\bar{s} = 1\}$$

where $n = \dim V$. Then by Hilbert 90, we have an isomorphism of F-varieties: $Res_{E/F}GL_n/GL_{n,F} \simeq S_n$. Then we see that the quotient $G//H \simeq S_n//GL_{n-1,F}$ where $GL_{n-1,F}$ acts on $S_n \subset Res_{E/F}GL_n$ by conjugation. Then the ring of invariants on S_n is generated by

$$tr \wedge^i s, es^j e^*, \quad i = 1,2,...,n, j = 0,1,...,n-1$$

where $e = (0,...,0,1)$ and e^* the transpose of e. Similarly, s is regular semisimple if and only if $det(es^{i+j}e^*)_{0\le i,j\le n-1} \neq 0$. We similarly define a notion of matching orbits. One can show that this defines a bijection

$$S_n(F)_{rs}//GL_{n-1}(F) \simeq \coprod_{W\subset V} U(V)(F)_{rs}//U(W)(F)$$

where the disjoint union in RHS runs over all pair $W \subset V$ such that the orthogonal complement of W in V is isometric to the one dimensional hermitian space which represents one (in particular, the isomorphism class of V is determined by that of W).

4.1.2. *The fundamental lemma of Jacquet–Rallis.* We now recall the fundamental lemma appeared in the Jacquet–Rallis relative trace formulae. Let E/F be a unramified quadratic extension of non-archimedean local fields with odd residue characteristic. On the linear side, we have the local orbital integral for $\gamma \in S_n(F)$:

$$O(\gamma, 1_{S_n(\mathcal{O}_F)}) = \int_{GL_{n-1}(F)} 1_{S_n(\mathcal{O}_F)}(h^{-1}\gamma h)(-1)^{v(det(h))} dh.$$

We normalize the measure such that the maximal compact open subgroup $GL_{n-1}(\mathcal{O}_F)$ has volume one. In this case, we have two isomorphism classes of hermitian space, denoted by W, W' respectively so that W has a self-dual lattice denoted by λ. Let Eu be a one dimensional hermitian space such that the norm of u is equal to one. Then the direct sum $V := W \oplus Eu$ has a self-dual lattice denoted by Λ. Its stabilizer denoted by $U(\Lambda)$ is a hyperspecial maximal open subgroup of $U(V)$. We similarly consider an orbital integral

$$O(\delta, 1_{U(\Lambda)}) = \int_{U(W)(F)} 1_{U(\Lambda)}(h^{-1}\delta h) dh$$

where the Haar measure on $U(W)(F)$ is normalized such that the volume of $U(\lambda) \subset U(W)$ is one.

Then the fundamental lemma (FL, for short) essentially conjectured by Jacquet–Rallis is as follows.

THEOREM 4.8. *Let δ be an orbit matching γ. Then when the residue characteristic is large, we have*

$$O(\gamma, 1_{S_n(\mathcal{O}_F)}) = \begin{cases} \pm O(\delta, 1_{U(\Lambda)}), & \delta \in U(W \oplus Eu); \\ 0, & \delta \in U(W' \oplus Eu). \end{cases}$$

This was proved by Zhiwei Yun when F is of positive characteristic $p > n$ ([**Y**]). J. Gordan transferred the result of Yun to the characteristic zero case when the residue characteristic is large. Note that the ambiguity of the sign $\pm$ can also be made precise as in [**Y**]. We will give an equivalent formulation of the FL without this ambiguity.

5. Arithmetic fundamental lemma

Inspired by the early joint work [**YZZ**] on Gross–Zagier formula and the relative trace formulae of Jacquet–Rallis, the author proposed a strategy to attack the arithmetic version Gan–Gross–Prasad conjecture for unitary Shimura varieties ([**Z**]). We now present the arithmetic fundamental lemma (AFL, for short) which plays the role of the FL in the Jacquet–Rallis approach.

5.1. Unitary Rapoport–Zink space. The Rapoport–Zink space in our setting serves as a sort of "local Shimura variety". We only discuss the case $F = \mathbb{Q}_p$. For a general finite extension of $\mathbb{Q}_p$, a parallel theory should exist but has been written in the literature. We first recall briefly the definition the unitary Rapoport–Zink space. For more details we refer to [**VW**]. Let $W = W(\mathbb{F}_p)$ be the Witt ring of the algebraic closure $\mathbb{F}_p$ of the finite field of p-elements and let $\mathcal{N}_n$ be the formal scheme over W that classifies quadruple $(X, \lambda, \iota, \rho)$ over a scheme S (over which p is locally nilpotent) over W:

- X is a p-divisible group over S of relative dimension n.
- $\lambda : X \to X^\vee$ prime-to-p principal polarization.
- $\iota : \mathcal{O}_E \to End_S(X)$.
- $\rho : X \times_W \mathbb{F}_p \to \mathbb{X}_n$ is a quasi-isogeny of height zero.

with a Kottwitz signature $(n-1, 1)$-condition. The data should also be compatible with each other. Here $(\mathbb{X}_n, \lambda_0, \iota_0)$ is a supersingular *basic object* over $\mathbb{F}_p$. It is essentially the formal completion of similar global moduli variety along the supersingular locus.

EXAMPLE 5.1. When $n = 1$, the universal object $\mathcal{X}_1 \to \mathcal{N}_1$ is the canonical lifting of $\mathbb{X}_1$ (a supersingular p-divisible group of dimension one) studied by Gross ([**Gr86**]).

We list some basic properties of $\mathcal{N}_n$:

- $\mathcal{N}_n$ is formally smooth of relative dimension $n-1$ over $Spf\, W$, locally formally of finite type.
- The reduced scheme $\mathcal{N}_n^{red}$ is a scheme of dimension $[\frac{n-1}{2}]$ over $\mathbb{F}_p$.

The proof can be found in [**RZ**] and [**VW**]. What is nice about the signature-$(n-1, 1)$ case is that $\mathcal{N}_n^{red}$ has a *Bruhat-Tits stratification* following Vollaard-Wedhorn ([**VW**]). Let V' be the n-dimensional hermitian space of *odd* determinant. Consider the set of vertices

$$\mathrm{Vert} := \{\Lambda \subset V' \text{ lattice} | p\,(\Lambda^*/\Lambda) = 0\}$$

where Λ^* is the hermitian dual of Λ.

Then we have a disjoint union

$$\mathcal{N}_n^{red} = \coprod_{\Lambda \in Vert} \mathcal{V}(\Lambda)^\circ$$

where $\mathcal{V}(\Lambda)^\circ$ is a certain Deligne–Lusztig variety associated to the unitary group of the reduction Λ^*/Λ viewed as a hermitian space over the residue field. The combinatorics of this stratification is controlled by the building of $SU(V')$.

On the formal scheme $\mathcal{N}_n$ there is a natural action by the automorphism group

$$G'_n := \{g \in \mathrm{End}^0_E(\mathbb{X}_n) \mid \quad gg^\dagger = 1\}.$$

Here $\dagger$ is the Rosati involution induced by the polarization λ. The action of $g \in G'$ on $\mathcal{N}_n$ by changing ρ. The group G'_n can be naturally identified with the unitary group $U(V')$.

5.2. Arithmetic Fundamental lemma. We now recall the statement of the arithmetic fundamental lemma [**Z**].

Fix an integer $n \geq 2$. We consider the product $\mathcal{N} = \mathcal{N}_{n-1} \times_{\mathrm{Spf}\, W} \mathcal{N}_n$. Note that there is a natural embedding $\mathcal{N}_{n-1} \to \mathcal{N}_n$. This induces a diagonal embedding $\Delta : \mathcal{N}_{n-1} \to \mathcal{N}$ whose image is denoted by $\Delta_{\mathcal{N}_{n-1}}$. We let W' be the $(n-1)$-dimensional hermitian space of *odd* determinant. Then we can write $V' = W' \oplus Eu$ for a vector u of norm one. Then $U(W') \times U(V')$ acts on $\mathcal{N}$ so we may consider the intersection for $g \in U(V')$:

$$(\Delta_{\mathcal{N}_{n-1}}, (1, g)\Delta_{\mathcal{N}_{n-1}})_{\mathcal{N}}.$$

On the other hand we may modify the orbital integral of Jacquet–Rallis as follows. We define for $s \in \mathbb{C}$ and a regular semisimple $\gamma \in S_n(F)_{rs}$:

$$O(\gamma, 1_{S_n(\mathcal{O}_F}, s) = \int_{GL_{n-1}(F)} 1_{S_n(\mathbb{O}_F)}(h^{-1}\gamma h)(-1)^{v(det(h))}|det(h)|^s dh.$$

It is a polynomial of $q^{\pm s}$ for where q is the cardinality of the residue of $\mathcal{O}_F$. Then by the FL in the previous section, when $\gamma \leftrightarrow g \in U(V')$, we have $O(\gamma, 1_{S_n(\mathcal{O}_F}, 0) = 0$. Hence we are lead to considering the first derivative $O(\gamma, 1_{S_n(\mathcal{O}_F}, s)$ at $s = 0$.

Then the arithmetic fundamental lemma as in [**Z**] can be stated as:

Conjecture 5.2. When $\gamma \leftrightarrow g \in U(V')$, we have

$$(\Delta_{\mathcal{N}_{n-1}}, (1, g)\Delta_{\mathcal{N}_{n-1}})_{\mathcal{N}} \cdot \log q = \pm O'(\gamma, 1_{S_n(\mathbb{Z}_p)}, 0).$$

Again the ambiguity of the sign $\pm$ can be made precise.

Usually such orbital integrals can be interpreted as counting the number of certain lattices. We now give such an interpretation for the AFL, at least in the best case of intersection.

We start with the fundamental lemma. We consider the following data:

- W: $(n-1)$-dimensional hermitian space of *even* determinant, $V = W \oplus Eu$, $(u, u) = 1$.
- Let $g \in U(V)$ and denote by $L = \sum_{i\geq 0} \mathcal{O}_E \cdot g^i u$. The regular semisimplicity of g is equivalent to that L is a lattice in V.
- $\mathrm{Fix}_L^g := \{\Lambda | L \subset \Lambda \subset L^*, g\Lambda = \Lambda\}$.
- Define an E-anti-linear involution τ on V characterized by $\tau(g^i u) = g^{-i}u$ for $i = 0, 1, ..., n-1$.

Then the FL can be restated as:

$$\sum_{\Lambda \in \mathrm{Fix}_L^g, \Lambda^* = \Lambda} 1 = \sum_{\Lambda \in \mathrm{Fix}_L^g, \Lambda^\tau = \Lambda} (-1)^{\ell(\Lambda/L)}.$$

To state the AFL, we consider similar data replacing W by W'. Then at least in the case of proper intersection, the AFL is equivalent to

$$\sum_{\Lambda \in \mathrm{Fix}_L^g, p\Lambda^*/\Lambda = 0} \mathrm{mult}(\Lambda) = \sum_{\Lambda \in \mathrm{Fix}_L^g, \Lambda^\tau = \Lambda} (-1)^{\ell(\Lambda/L)} \ell(\Lambda/L)$$

where $\mathrm{mult}(\Lambda)$ is the intersection multiplicity along the strata $\mathcal{V}(\Lambda)^\circ$. When the scheme theoretical intersection of $\Delta_{\mathcal{N}_{n-1}}$ and $(1, g)\Delta_{\mathcal{N}_{n-1}}$ is artinian, we may define

$$\mathrm{mult}(\Lambda) := \sum_{x \in \mathcal{V}(\Lambda)^\circ(\mathbb{F})} \mathrm{length}(g|x)$$

where $\mathrm{length}(g|x)$ is a certain length of deformation of endomorphism g of p-divisible group corresponding to the point x. Note that the condition $p\Lambda^*/\Lambda = 0$ can be thought as some "almost self-duality" for the lattice Λ.

5.3. Evidence. At this moment, the AFL remains open except in some lower dimensional cases. The main result of [**Z**] is

Theorem 5.3. *When $n = 2, 3$, the arithmetic fundamental lemma holds.*

When $n = 2$, the AFL is essentially proved in the original paper of Gross–Zagier ([**GZ**]) using Gross's theory of canonical lifting. When $n = 3$, the proof is by explicit computation of both sides. Either side is by no mean easy. To calculate the intersection number, we have to use the previous work of Gross, Keating and Kudla–Rapoport ([**KR**]).

Remark 5.4. We may also formulate an AFL for Example 2.3. Moreover in this case, together with other ingredients, we may reprove the Gross–Zagier formula in the case $\chi = 1$ (χ as in sec. 3) . We refer to [**TYZ**] for more details.

References

[AG] Aizenbud, Avraham; Gourevitch, Dmitry. *Generalized Harish-Chandra descent, Gelfand pairs, and an Archimedean analog of Jacquet-Rallis's theorem.* With an appendix by the authors and Eitan Sayag. Duke Math. J. 149 (2009), no. 3, 509–567.

[GGP] Wee Teck Gan, Benedict H. Gross and Dipendra Prasad. *Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups* Preprint, http://www.math.harvard.edu/∼gross/preprints/

[GJR1] Ginzburg, David; Jiang, Dihua; Rallis, Stephen *On the nonvanishing of the central value of the Rankin-Selberg L-functions.* J. Amer. Math. Soc. 17 (2004), no. 3, 679C722

[GJR2] Ginzburg, David; Jiang, Dihua; Rallis, Stephen *On the nonvanishing of the central value of the Rankin-Selberg L-functions. II.* Automorphic representations, L-functions and applications: progress and prospects, 157C191

[Gr86] Gross, Benedict H. *On canonical and quasicanonical liftings.* Invent. Math. 84 (1986), no. 2, 321–326.

[Gr] B. Gross: *Heegner points and representation theory.* Heegner points and Rankin L-series, 37–65, Math. Sci. Res. Inst. Publ., 49, Cambridge Univ. Press, Cambridge, 2004.

[GZ] B. Gross; D. Zagier: *Heegner points and derivatives of L-series.* Invent. Math. 84 (1986), no. 2, 225–320.

[I] Ichino Atsushi. *Trilinear forms and the central values of triple product L-functions.* Duke Math. Journal.

[II] A. Ichino and T. Ikeda. *On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture,*.preprint. math01.sci.osaka-cu.ac.jp/ ichino/gp.pdf

[J] Hervé Jacquet, *Sur un résultat de Waldspurger*, Ann. Sci. Éc. Norm. Supér., IV. 19 (1986), no. 2, 185-229

[JPSS] Jacquet, H.; Piatetskii-Shapiro, I. I.; Shalika, J. A. *Rankin- Selberg convolutions.* Amer. J. Math. 105 (1983), no. 2, 367–464.

[JR] Jacquet, Herve; Rallis, S. *On the Gross-Prasad conjecture for unitary groups.* http://www.math.columbia.edu/∼hj/GrossPrasadFinal.pdf

[K] S. Kudla: *Central derivatives of Eisenstein series and height pairings.* Ann. of Math. (2) 146 (1997), no. 3, 545–646.

[KR] Stephen Kudla, Michael Rapoport. *Special cycles on unitary Shimura varieties I. unramified local theory.* arXiv:0804.0600

[RZ] Rapoport, M.; Zink, Th. *Period spaces for p-divisible groups.* Annals of Mathematics Studies, 141. Princeton University Press, Princeton, NJ, 1996.

[Sa] H. Saito: *On Tunnell's formula for characters of* GL(2). Compositio Math. 85 (1993), no. 1, 99–108.

[S] Y.Sakellaridis. *Spherical varieties and integral representations of L-functions.* Submitted, arXiv:0905.4245.

[SV] Y. Sakellaridis, A. Venkatesh. *Periods and harmonic analysis on spherical varieties.*

[TYZ] Ye Tian, Xinyi Yuan, Wei Zhang. *A relative trace formula proof of the Gross-Zagier formula*, In preparation.

[Tu] J. Tunnell: *Local ϵ-factors and characters of* GL(2). Amer. J. Math. 105 (1983), no. 6, 1277–1307.

[VW] I. Vollaard, T. Wedhorn, *The supersingular locus of the Shimura variety for GU(1,n- 1), II.*,arXiv:0804.1522v1

[Wa] J. Waldspurger: *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie.* Compositio Math. 54 (1985), no. 2, 173–242.

[Y] Zhiwei Yun. *The fundamental lemma of Jacquet-Rallis*, Duke Math. J. 156 (2011), No.2, 167-228.

[YZZ] X. Yuan, S. Zhang, W. Zhang: *Gross–Zagier formula for Shimura curves.* to appear.

[Zh1] S. Zhang: *Heights of Heegner points on Shimura curves.* Ann. of Math. (2) 153 (2001), no. 1, 27–147.

[Zh2] S. Zhang: *Gross-Zagier formula for* GL_2. Asian J. Math. 5 (2001), no. 2, 183–290.

[Zh3] S. Zhang: *Gross-Zagier formula for* GL(2). II. Heegner points and Rankin L-series, 191–214, Math. Sci. Res. Inst. Publ., 49, Cambridge Univ. Press, Cambridge, 2004.

[Zh4] S. Zhang: *Linear forms, algebraic cycles, and derivatives of L-series.* preprint.

[Z] W. Zhang: *On arithmetic fundamental lemmas.* submitted.

Department of Mathematics, FAS, Harvard University, One Oxford Street, Cambridge MA 02138 USA

E-mail address: `wzhang@math.harvard.edu`