The Gross-Zagier formula: a brief introduction

David Hansen

March 28, 2011

The point of this talk is to give enough background to state the Gross-Zagier formula, and describe its immediate applications. I will prove almost nothing; the goal here is for you to see the formula and all its ingredients precisely. Time permitting, I will make some comments on the proof, and on more recent generalizations.

Let \( \mathfrak{D} = \{ x + iy \in \mathbb{C}, y > 0 \} \) be the upper half-plane, and let \( \Gamma_0(N) = \left\{ \gamma \in \operatorname{SL}_2(\mathbb{Z}) \mid \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\} \) act on the upper half-plane by linear fractional transformations. We may form the quotient \( Y_0(N) = \Gamma_0(N) \backslash \mathfrak{D} \), a Riemann surface with finitely many cusps. Compactifying gives a curve \( X_0(N) \) which is in fact defined over \( \mathbb{Q} \); the map \( z \mapsto (j(z), j(Nz)) \in \mathbb{A}_{\mathbb{C}}^2 \) realizes \( Y_0(N) \) as a (highly singular) plane curve with \( \mathbb{Q} \)-coefficients. Over a general field \( k \) of characteristic zero, the \( k \)-points of the curve \( X_0(N) \) (away from the cusps) parametrize diagrams \( (\phi : E \rightarrow E') \) where \( E/k, E'/k \) are elliptic curves and \( \phi : E \rightarrow E' \) is a \( k \)-rational isogeny with \( \ker \phi \cong \mathbb{Z}/N\mathbb{Z} \) over \( E \). There is a canonical \( \mathbb{Q} \)-rational involution \( w_N : X_0(N) \rightarrow X_0(N) \) which sends the diagram \( (\phi : E \rightarrow E') \) to the diagram \( (\hat{\phi} : E' \rightarrow E) \).

Over \( \mathbb{C} \), elliptic curves are simply quotients \( \mathbb{C}/\Lambda \) for lattices \( \Lambda = \omega_1 \mathbb{Z} + \omega_2 \mathbb{Z} \subset \mathbb{C} \), \( \omega_1/\omega_2 \notin \mathbb{R} \); the Weierstrass \( \wp \)-function

\[
\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{v \in \Lambda \setminus \{0\}} \left( \frac{1}{(z-v)^2} - \frac{1}{v^2} \right)
\]

yields an explicit uniformization of the corresponding curve via the map \( z \mapsto (1 : \wp_\Lambda'(z) : \wp_\Lambda(z)) \in \mathbb{P}_\mathbb{C}^1 \). Dilating \( \omega_1 \) and \( \omega_2 \) by a common scalar \( \lambda \) yields an isomorphic curve, since \( \wp_{\lambda \Lambda}(z) = \lambda^{-2} \wp_\Lambda(\lambda^{-1}z) \) so we may rescale by \( \omega_1^{-1} \) and consider the lattices \( \Lambda = \mathbb{Z} + \tau \mathbb{Z} \), assuming without any loss that \( \operatorname{Im} \tau > 0 \). Finally, two distinct points \( \tau, \tau' \in \mathfrak{D} \) yield homothetic lattices if and only if one is a translate of the other by an element of \( \operatorname{SL}_2(\mathbb{Z}) \), so the space of elliptic curves over \( \mathbb{C} \) is simply the quotient \( \operatorname{SL}_2(\mathbb{Z})/\mathfrak{D} = X_0(1) \). The \( \mathbb{C} \)-points of the covering \( X_0(N) \rightarrow X_0(1) \) correspond to diagrams (pr : \( \mathbb{C}/\Lambda' \rightarrow \mathbb{C}/\Lambda \)) for lattices \( \Lambda' \subset \Lambda \) with \( [\Lambda : \Lambda'] = N \), and the covering map is just the forgetful map (pr : \( \mathbb{C}/(\mathbb{Z} + \tau \mathbb{Z}) \rightarrow \mathbb{C}/(\frac{1}{N} \mathbb{Z} + \tau \mathbb{Z})) \rightarrow \mathbb{C}/(\mathbb{Z} + \tau \mathbb{Z}) \), \( \tau \in \mathfrak{D} \). The involution \( w_N \) acts by \( w_N(\tau) = \frac{-1}{N\tau} \).

There is a canonical construction of algebraic points on \( X_0(N) \). Let \( d < 0 \) be a quadratic discriminant, and let \( K = \mathbb{Q}(\sqrt{d}) \) be an imaginary quadratic field with Hilbert class field \( H_K \); class field theory yields a canonical isomorphism \( \operatorname{Art}_K : \operatorname{Cl}(K) \cong \operatorname{Gal}(H_K/K) \) mapping [\( p \)] to \( \operatorname{Frob}_p \). Suppose furthermore that every prime dividing \( N \) is split in \( K \) (this is the ubiquitous Heegner hypothesis). Then we may find some \( n \subset \mathcal{O}_K \) with \( \mathcal{O}_K/n \cong \mathbb{Z}/N\mathbb{Z} \); there are \( 2^{\omega(N)} \) such \( n \)'s, where \( \omega(N) \) is the number of distinct prime divisors of \( N \). Then for any ideal \( a \subset \mathcal{O}_K \), the diagram

\[
\begin{array}{ccc}
\mathcal{O}_K/a & \xrightarrow{\cdot a} & \mathcal{O}_K/a \\
\| & | & | \\
\tilde{\mathcal{O}}_K/a & \xrightarrow{\cdot a} & \tilde{\mathcal{O}}_K/a \\
\| & | & | \\
\mathbb{Z}/N\mathbb{Z} & \xrightarrow{\cdot a} & \mathbb{Z}/N\mathbb{Z}
\end{array}
\]

...
(pr : C/a → C/n⁻¹a) gives a point on X₀(N)(C). Dilating a by anything in K* gives the same elliptic curve, so this construction only depends on the image of a in the ideal class group Cl(K) of K. Hence we get a map

\[ \gamma_n : Cl(K) \to X₀(N)(C) \]
\[ [a] \mapsto (pr : C/a → C/n⁻¹a). \]

These points are actually defined over H_K and satisfy the Galois-equivariance property Art_K(ρ) · \gamma_n([a]) = \gamma_n([pa]) for all p. These are the **Heegner points**. We can be even more explicit. When N = 1 the Heegner hypothesis is always satisfied, and we get the usual points \( \gamma(a) = \frac{-b+\sqrt{d}}{2a} \in X₀(1) \) with \(-a < b < a \) and \( b² - 4ac = d \) for some \( c \geq a \). The choices of \( n \) biject with the solutions \( \beta \mod 2N \) of \( 2β² = d \mod 4N \) (exercise), and given \( β \) there is a unique \( Γ₀(N) \)-orbit of \( γ(a) \) containing a point \( \frac{-B+\sqrt{d}}{2A} \) with \( N|A \) and \( B \equiv β \mod 2N \). This is \( γ_n(a) \).

Now, let E/Q be an elliptic curve of conductor N, say \( E : y² = x³ + ax + b \) for some \( a, b ∈ Z \); the conductor is just some integer dividing the discriminant \( Δ = 16(4a³ + 27b²) \), which measures bad reduction in a “slightly more refined way” than \( Δ \) does (e.g. it only depends on the isogeny class of E). The version of modularity which people prove is an isomorphism between two ℓ-adic Galois representations; more relevantly for us, modularity means there is a (unique) modular form \( f_E(z) = \sum_{n=1}^∞ a_E(n)e^{2πinz} \) of weight 2 and level \( N \), such that \( |E(F_p)| = p + 1 - a_E(p) \) for all primes \( p \). Set

\[ \|f_E\|² = \int_{Γ₀(N)\backslash H} y²|f_E(z)|²dμ = \int_{Γ₀(N)\backslash H} |f(z)|²dxdy \]

for later use; note that this is well-defined because \( f(γz) = (cz + d)² \) and \( \text{Im}(γz) = \frac{Imz}{|cz + d|²} \), and positive because \( f_E \) is holomorphic and thus can only vanish on a countable set. Being modular also implies there is a **modular parametrization** of E, a dominant morphism \( φ_E : X₀(N) → E \) defined over Q. This is very deep; it comes from the embedding \( X₀(N) → \text{Jac}(X₀(N)) \), a construction of Shimura which yields a modular elliptic curve \( E' \) as a quotient of \( \text{Jac}(X₀(N)) \) which is modular and with \( f_E = f_E' \), and Faltings's isogeny theorem. There are several choices of \( φ_E \), but it becomes unique if we demand that \( φ_E(∞) = 0 \) and \( φ_E^*(dω) = 2πicf_E(z)dz \) for some \( c > 0 \), where \( dω = \frac{dz}{2y} \) is a translation-invariant 1-form. In fact, \( φ_E \) is given under these stipulations explicitly via

\[ φ_E(z) = -2πic \int_0^∞ f_E(τ)dτ. \]

Now, remember we have those Heegner points \( γ_n(\mathfrak{a}) ∈ X₀(N)(H_K) \) parametrized by the ideal class group of K. Composing with the modular parametrization gives a point \( P_{[a],n} := φ_E(γ_n([a])) \in E(H_K) \). It turns out that changing \( n \) changes all the \( P_{[a],n} \)'s by either nothing or by inversion, so I will henceforth fix \( n \) permanently and drop it from my notation. Adding up over \([a]\) with respect to the group law gives a point

\[ P_K = \sum_{[a] \in Cl(K)} P_{[a]} \]

which is contained in \( E(K) \); indeed, for any \( σ ∈ \text{Gal}(H_K/K) \), the action \( P_{[a]} → σP_{[a]} = P_{\text{Art}_{H_K}^{-1}(σ)[a]} \) simply permutes the ideal classes in the summation. The Gross-Zagier theorem describes the height of this point, in terms of an L-function.
The L-function of $E/\mathbb{Q}$ is
\[
L(s, E/\mathbb{Q}) := \prod_{p \mid N} \frac{1}{1 - a_E(p)p^{-s} + p^{1-2s}} \prod_{p \mid N} \frac{1}{1 - a_E(p)p^{-s}} = \sum_{n=1}^{\infty} a_E(n)n^{-s}.
\]

Modularity implies that this is holomorphic and that $\Lambda(s, E/\mathbb{Q}) := (2\pi)^{-s}N^{s/2}\Gamma(s)L(s, E/\mathbb{Q})$ satisfies

\[
\Lambda(s, E/\mathbb{Q}) = \pm \Lambda(2-s, E/\mathbb{Q}).
\]
This $\pm 1$ is the root number $\varepsilon(E/\mathbb{Q})$. More generally, given $K/\mathbb{Q}$ as before, there is a unique quadratic Dirichlet character $\chi_d$ of period $|d|$ with $\zeta_K(s) = \zeta_{\mathbb{Q}}(s)L(s, \chi_d)$, and we define the twisted L-function
\[
L(s, E_d/\mathbb{Q}) = \sum_{n=1}^{\infty} a_E(n)\chi_d(n)n^{-s} = \prod_{p \mid N} \frac{1}{1 - a_E(p)\chi_d(p)p^{-s} + \chi_d(p)^2 p^{1-2s}} \prod_{p \mid N} \frac{1}{1 - a_E(p)\chi_d(p)p^{-s}}.
\]

The notation is justified by the fact that this is the L-function of the curve $E_d : dy^2 = x^3 + ax + b$. This satisfies the same functional equation, with $N$ replaced by $Nd^2$, but with a different root number, namely $\varepsilon(E_d/\mathbb{Q}) = \varepsilon(E/\mathbb{Q})\chi_d(-N)$. Now set
\[
L(s, E/K) := L(s, E/\mathbb{Q})L(s, E_d/\mathbb{Q}).
\]

The notation is again justified by the fact that
\[
L(s, E/K) = \prod_{p \mid \mathcal{O}_K \cup \mathcal{N}_{\operatorname{disc}K}} \frac{1}{1 - a_E(p)\mathcal{N}p^{-s} + \mathcal{N}p^{1-2s}} \prod_{p \mid \mathcal{N}_{\operatorname{disc}K}} \ldots,
\]
where $a_E(p) = \mathcal{N}p + 1 - |E(\mathcal{O}_K/p)|$. What is the root number of this L-function? We compute
\[
\varepsilon(E/\mathbb{Q})\varepsilon(E_d/\mathbb{Q}) = \varepsilon(E/\mathbb{Q})^2\chi_d(-N) = \chi_d(-N) = -1
\]
since $d < 0$ and all the primes dividing $N$ are split in $K$. This forces $L(1, E/K) = 0$, and the Gross-Zagier formula computes $L'(1, E_K)$ as the value of a height function.

Given a finite extension $k/\mathbb{Q}$, let $M_k$ be the set of all places of $k$ and let $|\cdot|_v$ be the corresponding normalized valuation on $k_v$, i.e. $|x|_v = q_v^{-\operatorname{val}_v(x)}$ where $q_v$ is the cardinality of the residue field of $k_v$, and $\operatorname{val}_v(\varpi_v) = 1$ on a uniformizer. We have the product formula $\prod_{v \in M_k} |x|_v = 1 \forall x \in k$. For a point $x = (x_0 : x_1 : x_2) \in \mathbb{P}^2(k)$, define the height
\[
h_k(x) = \frac{1}{[k : \mathbb{Q}]} \log \left( \prod_{v \in M_k} \max \{|x_0|_v, |x_1|_v, |x_2|_v\} \right).
\]

Note that this is well-defined on projective space (by the product formula) and is nonnegative; the second property follows from $\prod_i \max \{a_i, b_i, \ldots\} \geq \max (\prod_i a_i, \prod_i b_i, \ldots)$ and the product formula. Note also that $h_E(x) = h_k(x)$ if $k \subset k'$, so the “direct limit”
\[
h(x) = \lim_{k} h_k(x)
\]
In particular, from this. Let’s start with the best one. Neron and Tate showed that this limit is well-defined, that $h_E(P)$ is a quadratic form on $E(K)$, and that $h_E(P) = 0$ if and only if $P \in E(K)_{\text{tors}}$.

**Theorem (Gross and Zagier).** With the above notation and assumptions, we have

$$L'(1, E_K) = \frac{32\pi^2 \|f_E\|^2}{|\phi_E|_K^2 |d| \deg \phi_E} h_E(P_K).$$

In particular,

$$L'(1, E_K) = 0 \iff P_K \text{ is torsion in } E(K).$$

(Note that $\frac{h_E(z)}{\deg \phi_E}$ is an isogeny invariant.) Gross and Zagier deduce several amazing corollaries from this. Let’s start with the best one.

**Corollary A.** If $E/Q$ is an elliptic curve with root number $\varepsilon = \varepsilon(E/Q) = -1$, and $L'(1, E/Q) \neq 0$, then $E(Q)$ contains elements of infinite order.

**Proof sketch.** This is not explained very well anywhere, so let me try. First, by a deep theorem of Waldspurger, we may find some $K$ satisfying the Heegner hypothesis with $L(1, E'_Q) \neq 0$.

We have $L'(1, E_K) = L'(1, E/Q)L(1, E'_Q) \neq 0$, so $P_K \in E(K)$ is nontorsion. Next, we need to understand the action of complex conjugation on the individual $P_n$’s. We shall do this by using the relations $w_N \cdot \gamma_n(a) = \gamma_n(-a^{-1})$ and $\gamma_n(a) = \gamma_n(\overline{a})$ on $X_0(N)$ together with the following lemma.

**Lemma A.1.** If $E/Q$ has root number $\varepsilon$, then for any $z \in X_0(N)(\mathbb{C})$, the point $\phi_E(z) + \varepsilon \phi_E(w_N \cdot z)$ is independent of $z$, and is torsion in $E(\mathbb{C})$.

**Proof.** Let $f = f_E$ be the newform corresponding to $E$, and write $\omega_f = 2\pi i c f(z)dz$ where $c$ is the Manin constant. By the Manin-Drinfeld theorem, the point

$$\phi_E(0) = -\int_0^{\infty} \omega_f$$

is torsion. On the other hand, we compute

$$\int_0^{\infty} \omega_f = \int_z^{\infty} \omega_f + \int_0^z \omega_f = \int_z^{\infty} \omega_f + \int_{w_Nz}^{\infty} w_N \omega_f = \int_z^{\infty} \omega_f - \int_{w_Nz}^{\infty} w_N \omega_f.$$

By newform theory, we know that $f(-1/Nz) = -\varepsilon z^2 Nf(z)$, and $d(-1/Nz)/dz = N^{-1}z^{-2}$, so $w_N \omega_f = -\varepsilon \omega_f$. Thus $\phi_E(0) = \phi_E(z) + \varepsilon \phi_E(w_Nz)$ for all $z \in X_0(N)(\mathbb{C})$. 


Applying the lemma with $z = \gamma\pi(a)$, and noting further that $\phi_E(z) = \phi_E(\overline{z})$, we compute
\[
\text{tors.} = \phi_E(\gamma(a)) + \phi_E(w_N \cdot \gamma(a)) = \phi_E(\gamma(a)) + \phi_E(\gamma(an^{-1})) = P[a] + \phi_{E}[a^{-1}n] = P[a] + \phi_{E}[\alpha^{-2}n] \\
\]
Since $\alpha$ was arbitrary, we conclude that if $\overline{\tau} \in \text{Gal}(H_K/Q) = \text{Gal}(H_K/K) \times \text{Gal}(K/Q)$ acts on $K$ nontrivially, then for any fixed $\alpha$, there is some $\sigma \in \text{Gal}(H/K)$ (depending on $\alpha$ and $\overline{\tau}$!) such that $\overline{\tau} P[a] + \varepsilon \sigma P[a]$ is torsion. Adding up the $\text{Gal}(H/K)$-translates of this, we find that
\[
\sum_{\rho \in \text{Gal}(H/K)} \rho \overline{\tau} P[a] + \varepsilon \rho \sigma P[a] = \sum_{\rho \in \text{Gal}(H/K)} \overline{\tau} P_{\text{Art}_K(\rho)[a]} + \varepsilon P_{\text{Art}_K(\sigma)[a]} = \overline{P} + \varepsilon P_K
\]
is torsion. By the parallelogram law for quadratic forms,
\[
h_{E}(\overline{P} - \varepsilon P_K) + h_{E}(\overline{P} + \varepsilon P_K) = 2h_{E}(P_K) + 2h_{E}(\overline{P}) = 4h_{E}(P_K) > 0
\]
so $\overline{P} - \varepsilon P_K \in E(K)$ is nontorsion and is defined over $Q$ iff $\varepsilon = -1$.

**Corollary B.** If $L(1, E/Q) \neq 0$ and $P_K$ is torsion for some $K$, then $L(s, E^d/Q)$ vanishes to order at least 3 at $s = 1$.

For example, this happens for the curve $E : y^2 = x^3 + 10x^2 - 20x + 8$ (of conductor 37) and $d = -139$. In this particular case, $E^{-139}$ provably has algebraic rank 3, and $L(s, E/K)$ provably vanishes to order exactly 3 at $s = 1$.

**Corollary C (Goldfeld).** There is an effective, computable constant $c > 0$ such that the class number of an imaginary quadratic field $Q(\sqrt{-d})$ satisfies
\[
|\text{Cl}(Q(\sqrt{-d}))| > c \log d \cdot \exp(-21 \sqrt{\log \log d}) >_{\varepsilon} (\log d)^{1-\varepsilon}.
\]

By the way, how did Heegner’s name get attached to these points? He used a proto-version of them to show, among other things, that the curve
\[
py^2 = x^3 - x
\]
has rational points of infinite order when $p$ is a prime with $p \equiv 5$ or 7 mod 8.