COMPUTATIONAL NUMBER THEORY IN RELATION WITH L-FUNCTIONS

HENRI COHEN
UNIVERSITÉ DE BORDEAUX,
INSTITUT DE MATHÉMATIQUES DE BORDEAUX,
351 COURS DE LA LIBÉRATION, 33405 TALENCE CEDEX, FRANCE

Abstract. We give a number of theoretical and practical methods related to the computation of L-functions, both in the local case (counting points on varieties over finite fields, involving in particular a detailed study of Gauss and Jacobi sums), and in the global case (for instance Dirichlet L-functions, involving in particular the study of inverse Mellin transforms); we also giving a number of little-known but very useful numerical methods, usually but not always related to the computation of L-functions.

1. L-Functions

This course is divided into four chapters. In the present first chapter, we introduce the notion of L-function, give a number of results and conjectures concerning them, and explain some of the computational problems in this theory. In the second chapter, we give a number of computational methods for obtaining the Dirichlet series coefficients of the L-function, so is *arithmetic* in nature. In the third chapter, we give a number of *analytic* tools necessary for working with L-functions. In the fourth and final chapter, we give a number of very useful numerical methods which are not sufficiently well-known, most of which being also related to the computation of L-functions.

1.1. **Introduction.** The theory of L-functions is one of the most exciting subjects in number theory. It includes for instance two of the crowning achievements of twentieth century mathematics, first the proof of the Weil conjectures and of the Ramanujan conjecture by Deligne in the early 1970's, using the extensive development of modern algebraic geometry initiated by Weil himself and pursued by Grothendieck and followers in the famous EGA and SGA treatises, and second the proof of the Shimura–Taniyama–Weil conjecture by Wiles et al., implying among other things the proof of Fermat's last theorem. It also includes two of the seven 1 million dollar Clay problems for the twenty-first century, first the Riemann hypothesis, and second the Birch–Swinnerton-Dyer conjecture which in my opinion is the most beautiful, if not the most important, conjecture in number theory, or even in the whole of mathematics.

Before giving a relatively general definition of L-functions, we look in some detail at a large number of special cases.

1.2. **The Prototype: the Riemann Zeta Function $\zeta(s)$.** The simplest of all (global) L-function is the Riemann zeta function $\zeta(s)$ defined by

$$\zeta(s)=\sum_{n\geq 1}\frac{1}{n^s}\;.$$

This is an example of a *Dirichlet series* (more generally $\sum_{n\geq 1}a(n)/n^s$, or even more generally $\sum_{n\geq 1}1/\lambda_n^s$, but we will not consider the latter). As such, it has a half-plane of absolute convergence, here $\Re(s)>1$.

The properties of this function, essentially studied initially by Bernoulli and Euler, are as follows, given historically:

(1) (Bernoulli, Euler): it has *special values*. When $s=2$, 4,... is a strictly positive even integer, $\zeta(s)$ is equal to π^s times a *rational number*. π is here a *period*, and is of course the usual π used for measuring circles. The rational numbers have *generating functions*, and are equal up to easy terms to the so-called *Bernoulli numbers*. For example $\zeta(2)=\pi^2/6$, $\zeta(4)=\pi^4/90$, etc... This was conjectured by Bernoulli and proved by Euler. Note that the proof in 1735 of the so-called *Basel problem*:

$$\zeta(2)=1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{\pi^2}{6}$$

is one of the crowning achievements of mathematics of that time.

(2) (Euler): it has an *Euler product*: for $\Re(s)>1$ one has the identity

$$\zeta(s)=\prod_{p\in P}\frac{1}{1-1/p^s}\;,$$

where P is the set of prime numbers. This is exactly equivalent to the so-called fundamental theorem of arithmetic. Note in passing (this does not seem interesting here but will be important later) that if we consider $1-1/p^s$ as a polynomial in $1/p^s=T$, its reciprocal roots all have the same modulus, here 1, this is of course trivial.

(3) (Riemann, but already "guessed" by Euler in special cases): it has an *analytic continuation* to a meromorphic function in the whole complex plane, with a single pole, at $s=1$, with residue 1, and a *functional equation* $\Lambda(1-s)=\Lambda(s)$, where $\Lambda(s)=\Gamma_{\mathbb{R}}(s)\zeta(s)$, with $\Gamma_{\mathbb{R}}(s)=\pi^{-s/2}\Gamma(s/2)$, and Γ is the gamma function.

(4) As a consequence of the functional equation, we have $\zeta(s)=0$ when $s=-2$, -4,..., $\zeta(0)=-1/2$, but we also have special values at $s=-1$, $s=-3$,... which are symmetrical to those at $s=2$, 4,... (for instance $\zeta(-1)=-1/12$, $\zeta(-3)=1/120$, etc...). This is the part which was guessed by Euler.

Roughly speaking, one can say that a global L-function is a function having properties similar to *all* the above. We will of course be completely precise below. Two things should be added immediately: first, the existence of special values will not be part of the definition but, at least conjecturally, a consequence. Second, all the global L-functions that we will consider should *conjecturally* satisfy a Riemann hypothesis: when suitably normalized, and excluding "trivial" zeros, all the zeros of the function should be on the line $\Re(s)=1/2$, axis of symmetry of the functional equation. Note that even for the simplest L-function, $\zeta(s)$, this is not proved.

1.3. **Dedekind Zeta Functions.** The Riemann zeta function is perhaps too simple an example to get the correct feeling about global L-functions, so we generalize:

Let K be a number field (a finite extension of $\mathbb{Q}$) of degree d. We can define its *Dedekind zeta function* $\zeta_K(s)$ for $\Re(s)>1$ by

$$\zeta_K(s)=\sum_{\mathfrak{a}}\frac{1}{\mathcal{N}(\mathfrak{a})^s}=\sum_{n\ge1}\frac{i(n)}{n^s}\ ,$$

where $\mathfrak{a}$ ranges over all (nonzero) integral ideals of the ring of integers $\mathbb{Z}_K$ of K, $\mathcal{N}(\mathfrak{a})=[\mathbb{Z}_K:\mathfrak{a}]$ is the norm of $\mathfrak{a}$, and $i(n)$ denotes the number of integral ideals of norm n.

This function has very similar properties to those of $\zeta(s)$ (which is the special case $K=\mathbb{Q}$). We give them in a more logical order:

(1) It can be analytically continued to the whole complex plane into a meromorphic function having a single pole, at $s=1$, with known residue, and it has a functional equation $\Lambda_K(1-s)=\Lambda_K(s)$, where

$$\Lambda_K(s)=|D_K|^{s/2}\Gamma_{\mathbb{R}}(s)^{r_1+r_2}\Gamma_{\mathbb{R}}(s+1)^{r_2}\ ,$$

where $(r_1,2r_2)$ are the number of real and complex embeddings of K and D_K its discriminant.

(2) It has an Euler product $\zeta_K(s)=\prod_{\mathfrak{p}}1/(1-1/\mathcal{N}(\mathfrak{p})^s)$, where the product is over all prime ideals of $\mathbb{Z}_K$. Note that this can also be written

$$\zeta_K(s)=\prod_{p\in P}\prod_{\mathfrak{p}|p}\frac{1}{1-1/p^{f(\mathfrak{p}/p)s}}\ ,$$

where $f(\mathfrak{p}/p)=[\mathbb{Z}_K/\mathfrak{p}:\mathbb{Z}/p\mathbb{Z}]$ is the so-called *residual index* of $\mathfrak{p}$ above p. Once again, note that if we set as usual $1/p^s=T$, the reciprocal roots of $1-T^{f(\mathfrak{p}/p)}$ all have modulus 1.

(3) It has *special values*, but only when K is a *totally real* number field ($r_2=0$, $r_1=d$): in that case $\zeta_K(s)$ is a *rational number* if s is a negative odd integer, or equivalently by the functional equation, it is a rational multiple of π^{ds} if s is a positive even integer.

An important new phenomenon occurs: recall that $\sum_{\mathfrak{p}|p}e(\mathfrak{p}/p)f(\mathfrak{p}/p)=d$, where $e(\mathfrak{p}/p)$ is the so-called *ramification index*, which is equivalent to the defining equality $p\mathbb{Z}_K=\prod_{\mathfrak{p}|p}\mathfrak{p}^{e(\mathfrak{p}/p)}$. In particular $\sum_{\mathfrak{p}|p}f(\mathfrak{p}/p)=d$ if and only if $e(\mathfrak{p}/p)=1$ for all $\mathfrak{p}$, which means that p is *unramified* in $K/\mathbb{Q}$; one can prove that this is equivalent to $p\nmid D_K$. Thus, the *local L-function* $L_{K,p}(T)=\prod_{\mathfrak{p}|p}(1-T^{f(\mathfrak{p}/p)})$ has degree in T exactly equal to d for all but a finite number of primes p, which are exactly those which divide the discriminant D_K, and for those "bad" primes the degree is strictly less than d. In addition, note that the number of $\Gamma_{\mathbb{R}}$ factors in the *completed* function $\Lambda_K(s)$ is equal to r_1+2r_2, hence once again equal to d.

Examples:

(1) The field $K=\mathbb{Q}(\sqrt{D})$ is a quadratic field of discriminant D. In that case, $\zeta_K(s)$ *factors* as $\zeta_K(s)=\zeta(s)L(\chi_D,s)$, where $\chi_D=\left(\dfrac{D}{.}\right)$ is the Legendre–Kronecker symbol, and $L(\chi_D,s)=\sum_{n\ge1}\chi_D(n)/n^s$. Thus, the local L-function at a prime p is given by

$$L_{K,p}(T)=(1-T)(1-\chi_D(p)T)=1-a_pT+\chi_D(p)T^2\ ,$$

with $a_p=1+\chi_D(p)$. Note that a_p is equal to the number of solutions in $\mathbb{F}_p$ of the equation $x^2=D$.

(2) Let us consider two special cases of (1): first $K=\mathbb{Q}(\sqrt{5})$. Since it is a real quadratic field, it has special values, for instance

$$\zeta_K(-1)=\frac{1}{30},\ \zeta_K(-3)=\frac{1}{60}\ ,\ \zeta_K(2)=\frac{2\sqrt{5}\pi^4}{375},\ \zeta_K(4)=\frac{4\sqrt{5}\pi^8}{84375}\ .$$

In addition, note that its *gamma factor* is $5^{s/2}\Gamma_{\mathbb{R}}(s)^2$.

Second, consider $K=\mathbb{Q}(\sqrt{-23})$. Since it is not a totally real field, $\zeta_K(s)$ does not have special values. However, because of the factorization $\zeta_K(s)=\zeta(s)L(\chi_D,s)$, we can look *separately* at the special values of $\zeta(s)$, which we have already seen (negative odd integers and positive even integers), and of $L(\chi_D,s)$. It is easy to prove that the special values of this latter function occurs at negative *even* integers and positive *odd* integers, which have empty intersection which those of $\zeta(s)$ and explains why $\zeta_K(s)$ itself has none. For instance,

$$L(\chi_D,-2)=-48\ ,\ L(\chi_D,-4)=6816\ ,L(\chi_D,3)=\frac{96\sqrt{23}\pi^3}{12167}\ .$$

In addition, note that its gamma factor is

$$23^{s/2}\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+1)=23^{s/2}\Gamma_{\mathbb{C}}(s)\ ,$$

where we set by definition

$$\Gamma_{\mathbb{C}}(s)=\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+1)=2\cdot(2\pi)^{-s}\Gamma(s)$$

by the duplication formula for the gamma function.

(3) Let K be the unique cubic field up to isomorphism of discriminant -23, defined for instance by a root of the equation $x^3-x-1=0$. We have $(r_1,2r_2)=(1,2)$ and $D_K=-23$. Here, one can prove (it is less trivial) that $\zeta_K(s)=\zeta(s)L(\rho,s)$, where $L(\rho,s)$ is a holomorphic function. Using both properties of ζ_K and ζ, this L-function has the following properties:

- It extends to an entire function on $\mathbb{C}$ with a functional equation $\Lambda(\rho,1-s)=\Lambda(\rho,s)$, with

$$\Lambda(\rho,s)=23^{s/2}\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+1)L(\rho,s)=23^{s/2}\Gamma_{\mathbb{C}}(s)L(\rho,s)\ .$$

  Note that this is the *same* gamma factor as for $\mathbb{Q}(\sqrt{-23})$. However the functions are fundamentally different, since $\zeta_{\mathbb{Q}(\sqrt{-23})}(s)$ has a pole at $s=1$, while $L(\rho,s)$ is an entire function.
- It is immediate to show that if we let $L_{\rho,p}(T)=L_{K,p}(T)/(1-T)$ be the local L function for $L(\rho,s)$, we have $L_{\rho,p}(T)=1-a_pT+\chi_{-23}(p)T^2$, with $a_p=1$ if $p=23$, $a_p=0$ if $\left(\frac{-23}{p}\right)=-1$, and $a_p=1$ or 2 if $\left(\frac{-23}{p}\right)=1$.

Remark: In all of the above examples, the function $\zeta_K(s)$ is *divisible* by the Riemann zeta function $\zeta(s)$, i.e., the function $\zeta_K(s)/\zeta(s)$ is an *entire function.* This is known for some number fields K, but is *not* known in general, even in degree $d=5$ for instance: it is a consequence of the more precise *Artin conjecture* on the holomorphy of Artin L-functions.

1.4. **Further Examples in Weight** 0. It is now time to give examples not coming from number fields. Define $a_1(n)$ by the formal equality

$$q\prod_{n\ge1}(1-q^n)(1-q^{23n})=\sum_{n\ge1}a_1(n)q^n\ ,$$

and set $L_1(s)=\sum_{n\ge1}a_1(n)/n^s$. The theory of modular forms (here of the Dedekind eta function) tells us that $L_1(s)$ will satisfy exactly the same properties as $L(\rho,s)$ with ρ as above.

Define $a_2(n)$ by the formal equality

$$\frac{1}{2}\left(\sum_{(m,n)\in\mathbb{Z}\times\mathbb{Z}} q^{m^2+mn+6n^2}-q^{2m^2+mn+3n^2}\right)=\sum_{n\geq 1}a_2(n)q^n\ ,$$

and set $L_2(s)=\sum_{n\geq 1}a_2(n)/n^s$. The theory of modular forms (here of theta functions) tells us that $L_2(s)$ will satisfy exactly the same properties as $L(\rho,s)$.

And indeed, it is an interesting *theorem* that

$$L_1(s)=L_2(s)=L(\rho,s)\ :$$

The "moral" of this story is the following, which can be made mathematically precise: if two L-functions are holomorphic, have the same gamma factor (including in this case the $23^{s/2}$), then they belong to a finite-dimensional vector space. Thus in particular if this vector space is 1-dimensional and the L-functions are suitably normalized (usually with $a(1)=1$), this implies as here that they are equal.

1.5. **Examples in Weight** 1**.** Although we have not yet defined the notion of weight, let me give two further examples.

Define $a_3(n)$ by the formal equality

$$q\prod_{n\geq 1}(1-q^n)^2(1-q^{11n})^2=\sum_{n\geq 1}a_3(n)q^n\ ,$$

and set $L_3(s)=\sum_{n\geq 1}a_3(n)/n^s$. The theory of modular forms (again of the Dedekind eta function) tells us that $L_3(s)$ will satisfy the following properties, analogous but more general than those satisfied by $L_1(s)=L_2(s)=L(\rho,s)$:

- It has an analytic continuation to the whole complex plane, and if we set

$$\Lambda_3(s)=11^{s/2}\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+1)L_3(s)=11^{s/2}\Gamma_{\mathbb{C}}(s)L_3(s)\ ,$$

  we have the functional equation $\Lambda_3(2-s)=\Lambda_3(s)$. Note the crucial difference that here $1-s$ is replaced by $2-s$.
- There exists an Euler product $L_3(s)=\prod_{p\in P}1/L_{3,p}(1/p^s)$ similar to the prededing ones in that $L_{3,p}(T)$ is for all but a finite number of p a second degree polynomial in T, and more precisely if $p=11$ we have $L_{3,p}(T)=1-T$, while for $p\neq 11$ we have $L_{3,p}(T)=1-a_pT+pT^2$, for some a_p such that $|a_p|<2\sqrt{p}$. This is expressed more vividly by saying that for $p\neq 11$ we have $L_{3,p}(T)=(1-\alpha_pT)(1-\beta_pT)$, where the reciprocal roots α_p and β_p have modulus exactly equal to $p^{1/2}$. Note again the crucial difference with "weight 0" in that the coefficient of T^2 is equal to p instead of ± 1, hence that $|\alpha_p|=|\beta_p|=p^{1/2}$ instead of 1.

As a second example, consider the equation $y^2+y=x^3-x^2-10x-20$ (an elliptic curve E), and denote by $N_q(E)$ the number of projective points of this curve over the finite field $\mathbb{F}_q$ (it is clear that there is a unique point at infinity, so if you want $N_q(E)$ is one plus the number of affine points). There is a universal recipe to construct an L-function out of a variety which we will recall below, but here let us simplify: for p prime, set $a_p = p+1-N_p(E)$ and

$$L_4(s) = \prod_{p\in P} 1/(1-a_pp^{-s}+\chi(p)p^{1-2s})\ ,$$

where $\chi(p)=1$ for $p\neq 11$ and $\chi(11)=0$. It is not difficult to show that $L_4(s)$ satisfies exactly the same properties as $L_3(s)$ (using for instance the elementary theory of modular curves), so by the moral explained above, it should not come as a surprise that in fact $L_3(s)=L_4(s)$.

1.6. **Definition of a Global L-Function.**

With all these examples at hand, it is quite natural to give the following definition of an L-function, which is not the most general but will be sufficient for us.

Definition 1.1. *Let d be a nonnegative integer. We say that a Dirichlet series $L(s)=\sum_{n\ge1}a(n)n^{-s}$ with $a(1)=1$ is an L-function of* degree *d and* weight 0 *if the following conditions are satisfied:*

(1) *(Ramanujan bound): we have $a(n)=O(n^{\varepsilon})$ for all $\varepsilon>0$, so that in particular the Dirichlet series converges absolutely and uniformly in any half plane $\Re(s)\ge\sigma>1$.*
(2) *(Meromorphy and Functional equation): The function $L(s)$ can be extended to $\mathbb{C}$ to a meromorphic function of order 1 having a finite number of poles; furthermore there exist complex numbers λ_i with nonnegative real part and an integer N called the* conductor *such that if we set*

$$\gamma(s) = N^{s/2}\prod_{1\le i\le d}\Gamma_{\mathbb{R}}(s+\lambda_i) \quad \textit{and} \quad \Lambda(s)=\gamma(s)L(s)\ ,$$

we have the functional equation

$$\Lambda(s)=\omega\overline{\Lambda(1-\overline{s})}$$

for some complex number ω, called the root number*, which will necessarily be of modulus 1.*
(3) *(Euler Product): For $\Re(s)>1$ we have an Euler product*

$$L(s)=\prod_{p\in P}1/L_p(1/p^s) \quad \textit{with} \quad L_p(T)=\prod_{1\le j\le d}(1-\alpha_{p,j}T)\ ,$$

and the reciprocal roots $\alpha_{p,j}$ are called the Satake parameters.
(4) *(Local Riemann hypothesis): for $p\nmid N$ we have $|\alpha_{p,j}|=1$, and for $p\mid N$ we have either $\alpha_{p,j}=0$ or $|\alpha_{p,j}|=p^{-m/2}$ for some m such that $1\le m\le d$.*

Remarks

1. More generally Selberg has defined a more general class of L-functions which first allows $\Gamma(\mu_i s + \lambda_i)$ with μ_i positive real in the gamma factors and second allows weaker assumptions on N and the Satake parameters.
2. Note that d is *both* the number of $\Gamma_{\mathbb{R}}$ factors, *and* the degree in T of the Euler factors $L_p(T)$, at least for $p \nmid N$, while the degree decreases for the "bad" primes p which divide N.
3. The Ramanujan bound (1) is easily seen to be a consequence of the conditions that we have imposed on the Satake parameters: in Selberg's more general definition this is not the case.

It is important to generalize this definition in the following trivial way:

Definition 1.2. *Let w be a nonnegative integer. A function $L(s)$ is said to be an L-function of degree d and* motivic weight w *if $L(s+(w-1)/2)$ is an L-function of degree d and weight* 0 *as above (with the slight additional technical condition that the nonzero Satake parameters $\alpha_{p,j}$ for $p \mid N$ satisfy $|\alpha_{p,j}| = p^{-m/2}$ with $1 \le m \le w$).*

For an L-function of weight w, it is clear that the functional equation is $\Lambda(s) = \omega\overline{\Lambda(k-\overline{s})}$ with $k = w+1$, and that the Satake parameters will satisfy $|\alpha_{p,j}| = p^{w/2}$ for $p \nmid N$, and for $p \mid N$ we have either $\alpha_{p,j} = 0$ or $|\alpha_{p,j}| = p^{(w-m)/2}$ for some integer m such that $1 \le m \le w$.

Thus, the first examples that we have given are all of weight 0, and the last two (which are in fact equal) are of weight 1. For those who know the theory of modular forms, note that the motivic weight (that we denote by w) is one less than the weight k of the modular form.

2. Origins of L-Functions

As can already be seen in the above examples, it is possible to construct L-functions in many different ways. In the present section, we look at three different ways for constructing L-functions: the first is by the theory of modular forms or more generally of *automorphic forms* (of which we have seen a few examples above), the second is by using Weil's construction of local L-functions attached to varieties and more generally to *motives*, and third, as a special but much simpler case of this, by the theory of *hypergeometric motives.*

2.1. L-Functions coming from Modular Forms.

The basic notion that we need here is that of *Mellin transform*: if $f(t)$ is a nice function tending to zero exponentially fast at infinity, we can define its Mellin transform $\Lambda(f;s) = \int_0^\infty t^s f(t)\,dt/t$, the integral being written in this way because dt/t is the invariant Haar measure on the locally compact group $\mathbb{R}_{>0}$. If we set $g(t) = t^{-k}f(1/t)$ and assume that g also tends to zero exponentially fast at infinity, it is immediate to see by a change of variable that $\Lambda(g;s) = \Lambda(f;k-s)$. This is exactly the type of functional equation needed for an L-function.

The other fundamental property of L-functions that we need is the existence of an Euler product of a specific type. This will come from the theory of *Hecke operators.*

A crash course in modular forms: we use the notation $q = e^{2\pi i\tau}$, for $\tau \in \mathbb{C}$ such that $\Im(\tau) > 0$, so that $|q| < 1$. A function $f(\tau) = \sum_{n\geq 1} a(n)q^n$ is said to be a modular cusp form of (positive, even) weight k if $f(-1/\tau) = \tau^k f(\tau)$ for all $\Im(\tau) > 0$. Note that because of the notation q we also have $f(\tau+1) = f(\tau)$, hence it is easy to deduce that $f((a\tau+b)/(c\tau+d)) = (c\tau+d)^k f(\tau)$ if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is an integer matrix of determinant 1. We define the L-function attached to f as $L(f;s) = \sum_{n\geq 1} a(n)/n^s$, and the Mellin transform $\Lambda(f;s)$ of the function $f(it)$ is on the one hand equal to $(2\pi)^{-s}\Gamma(s)L(f;s) = (1/2)\Gamma_{\mathbb{C}}(s)L(f;s)$, and on the other hand as we have seen above satisfies the functional equation $\Lambda(k-s) = (-1)^{k/2}\Lambda(s)$.

One can easily show the fundamental fact that the vector space of modular forms of given weight k is *finite dimensional*, and compute its dimension explicitly.

If $f(\tau) = \sum_{n\geq 1} a(n)q^n$ is a modular form and p is a prime number, one defines $T(p)(f)$ by $T(p)(f) = \sum_{n\geq 1} b(n)q^n$ with $b(n) = a(pn) + p^{k-1}a(n/p)$, where $a(n/p)$ is by convention 0 when $p \nmid n$, or equivalently

$$T(p)(f)(\tau) = p^{k-1}f(p\tau) + \frac{1}{p}\sum_{0\leq j<p} f\left(\frac{\tau+j}{p}\right)\ .$$

Then $T(p)f$ is also a modular cusp form, so $T(p)$ is an operator on the space of modular forms, and it is easy to show that the $T(p)$ commute and are diagonalizable, so they are simultaneously diagonalizable hence there exists a basis of common *eigenforms* for all the $T(p)$. Since one can show that for such an eigenform one has $a(1) \neq 0$, we can normalize them by asking that $a(1) = 1$, and we then obtain a canonical basis.

If $f(\tau) = \sum_{n\geq 1} a(n)q^n$ is such a *normalized eigenform*, it follows that the corresponding L function $\sum_{n\geq 1} a(n)/n^s$ will indeed have an Euler product, and using the elementary properties of the operators $T(p)$ that it will in fact be of the form:

$$L(f;s) = \prod_{p\in P} \frac{1}{1 - a(p)p^{-s} + p^{k-1-2s}}\ .$$

As a final remark, note that the analytic continuation and functional equation of this L-function is an *elementary consequence* of the definition of a modular form. This is totally different from the motivic cases that we will see below, where this analytic continuation is in general completely *conjectural.*

The above describes briefly the theory of modular forms on the modular group $\mathrm{PSL}_2(\mathbb{Z})$. One can generalize (nontrivially) this theory to *subgroups* of the modular group, the most important being $\Gamma_0(N)$ (matrices as above with $N \mid c$), to other *Fuchsian groups*, to forms in several variables, and even more generally to *reductive groups*.

2.2. **Local L-Functions of Algebraic Varieties.** Let V be some algebraic object. In modern terms, V may be a *motive*, whatever that may mean for the moment, but assume for instance that V is an algebraic variety, i.e., for each suitable field K, $V(K)$ is the set of common zeros of a family of polynomials in several variables. If K is a finite field $\mathbb{F}_q$ then $V(\mathbb{F}_q)$ will also be finite. For a number of easily explained reasons, one defines a local zeta function attached to V and a prime p (called the Hasse–Weil zeta function) as the formal power series in T

$$Z_p(V;T) = \exp\left(\sum_{n\ge1}\frac{|V(\mathbb{F}_{p^n})|}{n}T^n\right)\ .$$

There should be no difficulty in understanding this: setting for simplicity $v_n = |V(\mathbb{F}_{p^n})|$, we have

$$\begin{aligned}Z_p(V;T) &= \exp(v_1T + v_2T^2/2 + v_3T^3/3 + \cdots)\\ &= 1 + v_1T + (v_1^2 + v_2)T^2/2 + (v_1^3 + 3v_1v_2 + 2v_3)T^3/6 + \cdots\end{aligned}$$

For instance, if V is projective d-space P^d, we have $|V(\mathbb{F}_q)| = q^d + q^{d-1} + \cdots + 1$, and since $\sum_{n\ge1} p^{nj}T^n/n = -\log(1-p^jT)$, we deduce that $Z_p(P^d;T) = 1/((1-T)(1-pT)\cdots(1-p^dT))$.

After studying a number of special cases, such as elliptic curves (due to Hasse), and quasi-diagonal hypersurfaces in P^d, in 1949 Weil was led to make a number of conjectures on these zeta functions, assuming that V is a *smooth projective* variety, and proved these conjectures in the special case of curves (the proof is already quite deep).

The first conjecture says that $Z_p(V;T)$ is a *rational function* of T. This was proved by Dwork in 1960. Equivalently, this means that the sequence $v_n = |V(\mathbb{F}_{p^n})|$ satisfies a (non-homogeneous) linear recurrence with constant coefficients. For instance, if V is an *elliptic curve* defined over $\mathbb{Q}$ (such as $y^2 = x^3 + x + 1$) and if we set $a(p^n) = p^n + 1 - |V(\mathbb{F}_{p^n})|$, then

$$a(p^{n+1}) = a(p)a(p^n) - \chi(p)pa(p^{n-1})\ ,$$

where $\chi(p) = 1$ unless p divides the so-called *conductor* of the elliptic curve, in which case $\chi(p) = 0$ (this is not quite true because we must choose a suitable model for V, but it suffices for us).

The second conjecture of Weil states that this rational function is of the form

$$Z_p(V;T) = \prod_{0\le i\le 2d} P_{i,p}(V;T)^{(-1)^{i+1}} = \frac{P_{1,p}(V;T)\cdots P_{2d-1,p}(V;T)}{P_{0,p}(V;T)P_{2,p}(V;T)\cdots P_{2d,p}(V;T)}\ ,$$

where $d = \dim(V)$, and the $P_{i,p}$ are polynomials in T. Furthermore, Poincaré duality implies that $Z_p(V; 1/(p^d T)) = \pm p^{de/2} T^e Z_p(V; T)$ where e is the degree of the rational function (called the Euler characteristic of V), which means that there is a relation between $P_{i,p}$ and $P_{2d-i,p}$. In addition the $P_{i,p}$ have integer coefficients, and $P_{0,p}(T) = 1 - T$, $P_{2d,p}(T) = 1 - p^d T$. For instance, for *curves*, this means that $Z_p(V; T) = P_1(V; T)/((1 - T)(1 - pT))$, the polynomial P_1 is of even degree $2g$ (g is the genus of the curve) and satisfies $p^{dg} P_1(V; 1/(p^d T)) = \pm P_1(V; T)$.

For knowledgeable readers, in highbrow language, the polynomial $P_{i,p}$ is the reverse characteristic polynomial of the Frobenius endomorphism acting on the ith ℓ-adic cohomology group $H^i(V; \mathbb{Q}_\ell)$ for any $\ell \neq p$.

The third an most important of the Weil conjecture is the local *Riemann hypothesis*, which says that the reciprocal roots of $P_{i,p}$ have modulus exactly equal to $p^{i/2}$, i.e.,

$$P_{i,p}(V; T) = \prod_j (1 - \alpha_{i,j} T) \quad \text{with} \quad |\alpha_{i,j}| = p^{i/2} \ .$$

This last is the most important in applications.

The Weil conjectures were completely proved by Deligne in the early 1970's following a strategy already put forward by Weil, and is considered as one of the two or three major accomplishments of mathematics of the second half of the twentieth century.

Exercise: (You need to know some algebraic number theory for this). Let $P \in \mathbb{Z}[X]$ be a monic irreducible polynomial and $K = \mathbb{Q}(\theta)$, where θ is a root of P be the corresponding number field. Assume that $p^2 \nmid \operatorname{disc}(P)$. Show that the Hasse–Weil zeta function at p of the 0-dimensional variety defined by $P = 0$ is the Euler factor at p of the Dedekind zeta function $\zeta_K(s)$ attached to K, where p^{-s} is replaced by T.

2.3. **Global L-Function Attached to a Variety.** We are now ready to "globalize" the above construction, and build *global* L-functions attached to a variety.

Let V be an algebraic variety defined over $\mathbb{Q}$, say. We assume that V is "nice", meaning for instance that we choose V to be projective, smooth, and absolutely irreducible. For all but a finite number of primes p we can consider V as a smooth variety over $\mathbb{F}_p$, so for each i we can set $L_i(V; s) = \prod_p 1/P_{i,p}(V; p^{-s})$, where the product is over all the "good" primes, and the $P_{i,p}$ are as above. The factor $1/P_{i,p}(V; p^{-s})$ is as usual called the Euler factor at p. These functions L_i can be called the global L-functions attached to V.

This naïve definition is insufficient to construct interesting objects. First and most importantly, we have omitted a finite number of Euler factors at the so-called "bad primes", which include in particular those for which V is not smooth over $\mathbb{F}_p$, and although there do exist cohomological recipes to define them, as far as the author is aware these recipes do not really give practical algorithms. Another much less important reason is the fact that most of the L_i are uninteresting or related. For instance in the case of elliptic curves seen above, we have (up to a finite number of Euler factors) $L_0(V; s) = \zeta(s)$ and $L_2(V; s) = \zeta(s-1)$, so the only interesting L-function, called *the* L-function of the elliptic curve, is the function $L_1(V; s) = \prod_p (1 - a(p)p^{-s} + \chi(p)p^{1-2s})^{-1}$ (if the model of the curve is chosen to be *minimal*, this happens to be the correct definition, including for the "bad" primes). For varieties of higher dimension d, as we have mentioned as part of the Weil conjecture the functions L_i and L_{2d-i} are related by Poincaré duality, and L_0 and L_{2d} are translates of the Riemann zeta function (as above), so only the L_i for $1 \le i \le d$ need to be studied.

2.4. **Results and Conjectures on** $L(V; s)$**.** Global L-functions attached to varieties as above form a large source of L-functions: the problem with those functions is that most of their properties are only *conjectural*:

(1) The function L_i is only defined through its Euler product, and thanks to the last of Weil's conjectures, the local Riemann hypothesis, proved by Deligne, it converges absolutely for $\Re(s) > 1 + i/2$. Note that, with the definitions introduced above, L_i is an L-function of degree d_i, the common degree of $P_{i,p}$ for all but a finite number of p, and of weight exactly $w = i$ since the Satake parameters satisfy $|\alpha_{i,p}| = p^{i/2}$, again by the local Riemann hypothesis.

(2) A first conjecture is that L_i should have an *analytic continuation* to the whole complex plane with a *finite number* of *known* poles with *known* polar part.

(3) A second conjecture, which can in fact be considered as part of the first, is that this extended L-function should satisfy a *functional equation* when s is changed into $i + 1 - s$. More precisely, when completed with the Euler factors at the "bad" primes as mentioned (but not explained) above, then if we set

$$\Lambda_i(V; s) = N^{s/2} \prod_{1 \le j \le d_i} \Gamma_{\mathbb{R}}(s + \mu_i) L_i(V; s)$$

then $\Lambda_i(V; i+1-s) = \omega \overline{\Lambda_i(V^*; s)}$ for some variety V^* in some sense "dual" to V and a complex number ω of modulus 1. In the above, N is some integer divisible exactly by all the "bad" primes, i.e., essentially (but not exactly) the primes for which V reduced modulo p is not smooth, and the μ_i are in this case (varieties) *integers* which can be computed in terms of the *Hodge numbers* $h^{p,q}$ of the variety thanks to a recipe due to Serre.

In many cases the L-function is self-dual, in which case the functional equation is simply of the form $\Lambda_i(V; i+1-s) = \pm\Lambda_i(V; s)$.

(4) The function Λ_i should satisfy the generalized Riemann hypothesis (GRH): all its zeros in $\mathbb{C}$ are on the vertical line $\Re(s) = (i+1)/2$. Equivalently, the zeros of L_i are on the one hand real zeros at some integers coming from the poles of the gamma factors, and all the others satisfy $\Re(s) = (i+1)/2$.

(5) The function Λ_i should have *special values*: for the integer values of s (called special points) which are those for which neither the gamma factor at s nor at $i+1-s$ has a pole, it should be computable "explicitly": it should be equal to a *period* (integral of an algebraic function on an algebraic cycle) times an algebraic number. This has been stated (conjecturally) in great detail by Deligne in the 1970's.

It is conjectured that *all* L-functions of degree d_i and weight i as defined at the beginning should satisfy all the above properties, not only the L-functions coming from varieties.

I now list a number of cases where the above conjectures are proved.

(1) The first conjecture (analytic continuation) is known only for a very restricted class of L-functions: first L-functions of degree 1, which can be shown to be Dirichlet L-functions, L-functions attached to modular forms as shown above, and more generally to *automorphic forms*. For L-functions attached to varieties, one knows this *only* when one can prove that the corresponding L-function comes from an automorphic form: this is how Wiles proves the analytic continuation of the L-function attached to an elliptic curve, a very deep and difficult result, with Deligne's proof of the Weil conjectures one of the most important result of the end of the 20th century. More results of this type are known for certain higher-dimensional varieties such as certain *Calabi–Yau manifolds*, thanks to the work of Harris and others. Note however that for such simple objects as most *Artin L-functions* (degree 0) or Abelian surfaces, this is not known, although the work of Brumer–Kramer–Poor–Yuen on the *paramodular conjecture* may some day lead to a proof in this last case.

(2) The second conjecture on the existence of a functional equation is of course intimately linked to the first, and the work of Wiles et al. and Harris et al. also prove the existence of this functional equation. But in addition, in the case of Artin L-functions for which only meromorphy (possibly with infinitely many poles) is known thanks to a theorem of Brauer, this same theorem implies the functional equation which is thus known in this case. Also, as mentioned, the Euler factors which we must include for

the “bad” primes in order to have a clean functional equation are often quite difficult to compute.

(3) The Riemann hypothesis is not known for *any* global L-function of the type mentioned above, not even for the simplest one, the Riemann zeta function $\zeta(s)$. Note that it *is* known for other kinds of L-functions such as *Selberg zeta functions*, but these are functions of *order* 2 (growth at infinity like $e^{|s|^{2+\varepsilon}}$ instead of $e^{|s|^{1+\varepsilon}}$, so are not in the class considered above.

(4) Concerning *special values*: many cases are known, and many conjectured. This is probably one of the most *fun* conjectures since everything can be computed explicitly to hundreds of decimals if desired. For instance, for modular forms it is a theorem of Manin, for symmetric squares of modular forms it is a theorem of Rankin, and for higher symmetric powers one has very precise conjectures of Deligne, which check perfectly on a computer, but none of them are proved. For the Riemann zeta function or Dirichlet L-functions, of course all these results such as $\zeta(2) = \pi^2/6$ date back essentially to Euler.

In the case of an elliptic curve E over $\mathbb{Q}$, the only special point is $s = 1$, and in this case the whole subject evolves around the *Birch and Swinnerton-Dyer conjecture* which predicts the behavior of $L_1(E; s)$ around $s = 1$. The only known results, already quite deep, due to Kolyvagin and Gross–Zagier, deal with the case where the *rank* of the elliptic curve is 0 or 1.

There exist a number of other very important conjectures linked to the behavior of L-functions at integer points which are not necessarily special, such as the Bloch–Kato, Beilinson, Lichtenbaum, or Zagier conjecture, but it would carry us too far afield to describe them.

2.5. **Hypergeometric Motives.** Still another way to construct L-functions is through the use of *hypergeometric motives*, due to Katz and Rodriguez-Villegas. Although this construction is a special case of the construction of L-functions of varieties studied above, the corresponding variety is *hidden* (although it can be recovered if desired), and the computations are in some sense much simpler.

Let me give a short and unmotivated introduction to the subject: let $\gamma(T) = \sum_{n\geq 1} \gamma_n T^n \in \mathbb{Z}[T]$ be a polynomial satisfying the conditions $\gamma(0) = 0$ and $\gamma'(1) = 0$ (in other words $\gamma_0 = 0$ and $\sum_n n\gamma_n = 0$). For any finite field $\mathbb{F}_q$ with $q = p^f$ and any character χ of $\mathbb{F}_q^*$, recall that the Gauss sum $\mathfrak{g}(\chi)$ is defined by

$$\mathfrak{g}(\chi) = \sum_{x\in\mathbb{F}_q^*} \chi(x) \exp(2\pi i \operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(x)/p) \ .$$

We set

$$Q_q(\gamma;\chi)=\prod_{n\ge1}\mathfrak{g}(\chi^n)^{\gamma_n}$$

and for any $t\in\mathbb{F}_q\setminus\{0,1\}$

$$a_q(\gamma;t)=\frac{q^d}{1-q}\left(1+\sum_{\chi\ne\chi_0}\chi(Mt)Q_q(\gamma;\chi)\right)\ ,$$

where χ_0 is the trivial character, d is an integer, and M a nonzero rational number, both of which can easily be given explicitly (M is simply a normalization parameter, since one could change Mt into t). Then the "theorem" of Katz is that for $t\ne0,1$ the quantity $a_q(\gamma;t)$ is the *trace of Frobenius* on some *motive defined over* $\mathbb{Q}$ (I put theorem in quotes because it is not completely clear what the status of the proof is, although there is no doubt that it is true). In the language of L-functions this means the following: define as usual the local L-function at p by the formal power series

$$L_p(\gamma;t;T)=\exp\left(\sum_{f\ge1}a_{p^f}(\gamma;t)\frac{T^f}{f}\right)\ .$$

Then L_p is a rational function of T, satisfies the local Riemann hypothesis, and if we set

$$L(\gamma;t;s)=\prod_p L_p(\gamma;t;p^{-s})^{-1}\ ,$$

then L once completed at the "bad" primes should be a global L-function of the standard type described above.

2.6. **Computational Goals.** Now that we have a handle on what L-functions are, we come to the computational and algorithmic problems, which are the main focus of these notes. This involves many different aspects, all interesting in their own right.

In a first type of situation, we assume that we are "given" the L-function, in other words that we are given a reasonably "efficient" algorithm to compute the coefficients $a(n)$ of the Dirichlet series (or the Euler factors), and that we know the gamma factor $\gamma(s)$. The main computational goals are then the following:

(1) Compute $L(s)$ for "reasonable" values of s: for example, compute $\zeta(3)$. More sophisticated, but much more interesting: compute special values of symmetric powers L-functions of modular forms, and check numerically the conjectures of Deligne on the subject.

(2) Check the numerical validity of the functional equation, and in passing, if unknown, compute the numerical value of the *root number* ω occurring in the functional equation.

(3) Compute $L(s)$ for $s = 1/2 + it$ for rather large real values of t (in the case of weight 0, more generally for $s = (w+1)/2 + it$), and/or make a plot of the corresponding Z function (see below).
(4) Compute all the zeros of $L(s)$ on the critical line up to a given height, and check the corresponding Riemann hypothesis.
(5) Compute the residue of $L(s)$ at $s = 1$ (typically): for instance if L is the Dedekind zeta function of a number field, this gives the product hR.
(6) Compute the *order* of the zeros of $L(s)$ at integer points (if it has one), and the leading term in the Taylor expansion: for instance for the L-function of an elliptic curve and $s = 1$, this gives the *analytic rank* of an elliptic curve, together with the Birch and Swinnerton-Dyer data.

Unfortunately, we are not always given an L-function completely explicitly. We can lack more or less partial information on the L-function:

(1) One of the most frequent situations is that one knows the Euler factors for the "good" primes, as well as the corresponding part of the conductor, and that one is lacking both the Euler factors for the bad primes and the bad part of the conductor. The goal is then to find numerically the missing factors and missing parts.
(2) A more difficult but much more interesting problem is when essentially nothing is known on the L-function except $\gamma(s)$, in other words the $\Gamma_{\mathbb{R}}$ factors and the constant N, essentially equal to the conductor. It is quite amazing that nonetheless one can quite often tell whether an L-function with the given data can exist, and give some of the initial Dirichlet coefficients (even when several L-functions may be possible).
(3) Even more difficult is when essentially nothing is known except the degree d and the constant N, and one looks for possible $\Gamma_{\mathbb{R}}$ factors: this is the case in the search for Maass forms over $\mathrm{SL}_n(\mathbb{Z})$, which has been conducted very successfully for $n = 2$, 3, and 4.

For lack of time, we will not say much about these problems.

2.7. **Available Software for L-Functions.** Many people working on the subject have their own software. I mention the available public data.

• M. Rubinstein's `C++` program `lcalc`, which can compute values of L-functions, make large tables of zeros, and so on. The program uses `C++` language `double`, so is limited to 15 decimal digits, but is highly optimized, hence very fast, and used in most situations. Also optimized for large values of the imaginary part using Riemann–Siegel. Available in `Sage`.

- T. Dokshitser's program `computel`, initally written in `GP/Pari`, rewritten for `magma`, and also available in `Sage`. Similar to Rubinstein's, but allows arbitrary precision, hence slower, and has no built-in zero finder, although this is not too difficult to write. Not optimized for large imaginary parts.

- Last but not least, not a program but a huge *database* of L-functions, modular forms, number fields, etc..., which is the result of a collaborative effort of approximately 30 to 40 people headed by D. Farmer. This database can of course be queried in many different ways, it is possible and useful to navigate between related pages, and it also contains `knowls`, bits of knowledge which give the main definitions. In addition to the stored data, the site can compute on the fly (using the software mentioned above, essentially `Pari`, `Sage`, and `lcalc`) additional required information. Available at:
`http://www.lmfdb.org`

3. Arithmetic Methods: Computing $a(n)$

We now come to the second chapter of these notes: the computation of the Dirichlet series coefficients $a(n)$ and/or of the Euler factors, which is usually the same problem. Of course this depends entirely on how the L-function is *given*: in view of what we have seen, it can be given for instance (but not only) as the L-function attached to a modular form, to a variety, or to a hypergeometric motive. Since there are so many relations between these L-functions (we have seen several identities above), we will not separate the way in which they are given, but treat everything at once.

In view of the preceding section, an important computational problem is the computation of $|V(\mathbb{F}_q)|$ for a variety V. This may of course be done by a naïve point count: if V is defined by polynomials in n variables, we can range through the q^n possibilities for the n variables and count the number of common zeros. In other words, there always exists a trivial algorithm requiring q^n steps. We of course want something better.

3.1. **General Elliptic Curves.** Let us first look at the special case of *elliptic curves*, i.e., a projective curve V with affine equation $y^2 = x^3 + ax + b$ such that $p \nmid 6(4a^3 + 27b^2)$, which is almost the general equation for an *elliptic curve*. For simplicity assume that $q = p$, but it is immediate to generalize. If you know the definition of the Legendre symbol, you know that the number of solutions to $y^2 = n$ is equal to $1 + \left(\frac{n}{p}\right)$. Otherwise, since $\mathbb{F}_p$ is a field, it is clear that this number is equal to 0, 1, or 2, and so one can *define* $\left(\frac{n}{p}\right)$ as one less, so -1, 0, or 1. Thus, since it is immediate to see that there is a single projective point at infinity, we have

$$|V(\mathbb{F}_p)| = 1 + \sum_{x\in\mathbb{F}_p}\left(1+\left(\frac{x^3+ax+b}{p}\right)\right) = p+1-a(p)\ , \quad \text{with}$$

$$a(p) = -\sum_{0\le x\le p-1}\left(\frac{x^3+ax+b}{p}\right)\ .$$

Now a Legendre symbol can be computed very efficiently using the *quadratic reciprocity law*. Thus, considering that it can be computed in constant time (which is not quite true but almost), this gives a $O(p)$ algorithm for computing $a(p)$, already much faster than the trivial $O(p^2)$ algorithm consisting in looking at all pairs (x,y).

To do better, we have to use an additional and crucial property of an elliptic curve: it is an *abelian group*. Using this combined with the so-called Hasse bounds $|a(p)| < 2\sqrt{p}$ (a special case of the Weil conjectures), and the so-called *baby-step giant-step algorithm* due to Shanks, one can obtain a $O(p^{1/4})$ algorithm, which is very fast for all practical purposes.

However a remarkable discovery due to Schoof in the early 1980's is that there exists a practical algorithm for computing $a(p)$ which is *polynomial in* $\log(p)$, for instance $O(\log^6(p))$. Several important improvements have been made on this basic algorithm, in particular by Atkin and Elkies, and the resulting SEA algorithm (which is implemented in many computer packages) is able to compute $a(p)$ for p with several thousand decimal digits. Note however that in practical ranges (say $p < 10^{12}$), the $O(p^{1/4})$ algorithm mentioned above is sufficient.

3.2. **Elliptic Curves with Complex Multiplication.** In certain special cases it is possible to compute $|V(\mathbb{F}_q)|$ for an elliptic curve V much faster than with any of the above methods: when the elliptic curve V has *complex multiplication*. Let us consider the special cases $y^2 = x^3 - nx$ (the general case is more complicated but not really slower). By the general formula for $a(p)$, we have for $p \ge 3$:

$$\begin{aligned} a(p) &= -\sum_{-(p-1)/2\le x\le (p-1)/2}\left(\frac{x(x^2-n)}{p}\right)\\ &= -\sum_{1\le x\le (p-1)/2}\left(\left(\frac{x(x^2-n)}{p}\right)+\left(\frac{-x(x^2-n)}{p}\right)\right)\\ &= -\left(1+\left(\frac{-1}{p}\right)\right)\sum_{1\le x\le (p-1)/2}\left(\frac{x(x^2-n)}{p}\right)\end{aligned}$$

by the multiplicative property of the Legendre symbol. This already shows that if $\left(\frac{-1}{p}\right) = -1$, in other words $p \equiv 3 \pmod 4$, we have $a(p) = 0$. But we can also find a formula when $p \equiv 1 \pmod 4$: recall that in that case by a famous theorem due to Fermat, there exist integers u and v such that $p = u^2 + v^2$. If necessary by exchanging u and v, and/or changing the sign of u, we may assume that $u \equiv -1 \pmod 4$, in which case the decomposition is unique, up to the sign of v. It is then not difficult to prove the following theorem (see Section 8.5.2 of [4] for the proof):

Theorem 3.1. *Assume that* $p \equiv 1 \pmod 4$ *and* $p = u^2 + v^2$ *with* $u \equiv -1 \pmod 4$. *The number of projective points on the elliptic curve* $y^2 = x^3 - nx$ *(where* $p \nmid n$*) is equal to* $p + 1 - a(p)$*, where*

$$a(p) = 2\left(\frac{2}{p}\right)\begin{cases} -u & \text{if } n^{(p-1)/4} \equiv 1 \pmod p \\ u & \text{if } n^{(p-1)/4} \equiv -1 \pmod p \\ -v & \text{if } n^{(p-1)/4} \equiv -u/v \pmod p \\ v & \text{if } n^{(p-1)/4} \equiv u/v \pmod p \end{cases}$$

(note that one of these four cases must occur).

To apply this theorem from a computational standpoint we note the following two *facts*:

(1) The quantity $a^{(p-1)/4} \bmod p$ can be computed efficiently by the *binary powering algorithm* (in $O(\log^3(p))$ operations). It is however possible to compute it more efficiently in $O(\log^2(p))$ operations using the *quartic reciprocity law.*

(2) The numbers u and v such that $u^2 + v^2 = p$ can be computed efficiently (in $O(\log^2(p))$ operations) using *Cornacchia's algorithm* which is very easy to describe but not so easy to prove. It is a variant of Euclid's algorithm. It proceeds as follows:

- As a first step, we compute a square root of -1 modulo p, i.e., an x such that $x^2 \equiv -1 \pmod p$. This is done by choosing randomly a $z \in [1, p-1]$ and computing the Legendre symbol $\left(\frac{z}{p}\right)$ until it is equal to -1 (we can also simply try $z = 2$, 3, ...). Note that this is a fast computation. When this is the case, we have by definition $z^{(p-1)/2} \equiv -1 \pmod p$, hence $x^2 \equiv -1 \pmod p$ for $x = z^{(p-1)/4} \bmod p$. Reducing x modulo p and possibly changing x into $p - x$, we normalize x so that $p/2 < x < p$.

- As a second step, we perform the Euclidean algorithm on the pair (p, x), writing $a_0 = p$, $a_1 = x$, and $a_{n-1} = q_n a_n + a_{n+1}$ with $0 \leq a_{n+1} < a_n$, and we stop at the exact n for which $a_n^2 < p$. It can be proved (this is the difficult part) that for this specific n we have $a_n^2 + a_{n+1}^2 = p$, so up to exchange of u and v and/or change of signs, we can take $u = a_n$ and $v = a_{n+1}$.

Note that Cornacchia's algorithm can easily be generalized to solving efficiently $u^2 + dv^2 = p$ or $u^2 + dv^2 = 4p$ for any $d \geq 1$, see [2] and Cremona's course (incidentally one can also solve this for $d < 0$, but it poses completely different problems since there may be infinitely many solutions).

The above theorem is given for the special elliptic curves $y^2 = x^3 - nx$ which have complex multiplication by the (ring of integers of the) field $\mathbb{Q}(i)$, but a similar theorem is valid for all curves with complex multiplication, see Section 8.5.2 of [4].

3.3. **Using Modular Forms of Weight 2.**

By Wiles' celebrated theorem, the L-function of an elliptic curve is equal to the L-function of a modular form of weight 2 for $\Gamma_0(N)$, where N is the conductor of the curve. We do not need to give the precise definitions of these objects, but only a specific example.

Let V be the elliptic curve with affine equation $y^2+y = x^3-x^2$. It has conductor 11. It can be shown using classical modular form methods (i.e., without Wiles' theorem) that the global L-function $L(V; s) = \sum_{n\geq 1} a(n)/n^s$ is the same as that of the modular form of weight 2 over $\Gamma_0(11)$ given by

$$f(\tau) = q \prod_{m\geq 1} (1 - q^m)^2 (1 - q^{11m})^2 \ ,$$

with $q = \exp(2\pi i\tau)$. Even with no knowledge of modular forms, this simply means that if we formally expand the product on the right hand side as

$$q \prod_{m\geq 1} (1 - q^m)^2 (1 - q^{11m})^2 = \sum_{n\geq 1} b(n) q^n \ ,$$

we have $b(n) = a(n)$ for all n, and in particular for $n = p$ prime. We have already seen this example above with a slightly different equation for the elliptic curve (which makes no difference for its L-function outside of the primes 2 and 3).

We see that this gives an alternate method for computing $a(p)$ by expanding the infinite product. Indeed, the function

$$\eta(\tau) = q^{1/24} \prod_{m\geq 1} (1 - q^m)$$

is a modular form of weight 1/2 with known expansion:

$$\eta(\tau) = \sum_{n\geq 1} \left(\frac{12}{n}\right) q^{n^2/24} \ ,$$

and so using Fast Fourier Transform techniques for formal power series multiplication we can compute all the coefficients $a(n)$ simultaneously (as opposed to one by one) for $n \leq B$ in time $O(B \log^2(B))$. This amounts to computing each individual $a(n)$ in time $O(\log^2(n))$, so it seems to be competitive with the fast methods for elliptic curves with complex multiplication, but this is an illusion since we must store all B coefficients, so it can be used only for $B \leq 10^{12}$, say, far smaller than what can be reached using Schoof's algorithm, which is truly polynomial in $\log(p)$ for each fixed prime p.

3.4. **Higher Weight Modular Forms.** It is interesting to note that the dichotomy between elliptic curves with or without complex multiplication is also valid for modular forms of higher weight (again, whatever that means, you do not need to know the definitions). For instance, consider

$$\Delta(\tau) = \Delta_{24}(\tau) = \eta^{24}(\tau) = q\prod_{m\geq 1}(1-q^m)^{24} := \sum_{n\geq 1}\tau(n)q^n\ .$$

The function $\tau(n)$ is a famous function called the *Ramanujan τ function*, and has many important properties, analogous to those of the $a(p)$ attached to an elliptic curve (i.e., to a modular form of weight 2).

There are several methods to compute $\tau(p)$ for p prime, say. One is to do as above, using FFT techniques. The running time is similar, but again we are limited to $B \leq 10^{12}$, say. A second more sophisticated method is to use the *Eichler–Selberg trace formula*, which enables the computation of an individual $\tau(p)$ in time $O(p^{1/2+\varepsilon})$ for all $\varepsilon > 0$. A third very deep method, developed by Edixhoven, Couveignes, et al., is a generalization of Schoof's algorithm. While in principle polynomial time in $\log(p)$, it is not yet practical compared to the preceding method.

For those who want to see the formula explicitly, we let $H(N)$ be the *Hurwitz class number* (a small modification of the class number of imaginary quadratic fields), and $H_2(N) = H(N) + 2H(N/4)$, where we note that $H(N)$ can be computed in terms of $H(N/4)$ if the latter is nonzero. Then for p prime

$$\tau(p) = 28p^6 - 28p^5 - 90p^4 - 35p^3 - 1 - 128\sum_{1\leq t<p^{1/2}} t^6(4t^4 - 9pt^2 + 7p^2)H_2(4(p-t^2))\ ,$$

which is the fastest practical formula that I know for computing $\tau(p)$.

On the contrary, consider

$$\Delta_{26}(\tau) = \eta^{26}(\tau) = q^{13/12}\prod_{m\geq 1}(1-q^m)^{26} := q^{13/12}\sum_{n\geq 1}\tau_{26}(n)q^n\ .$$

This is what is called a modular form with complex multiplication. Whatever the definition, this means that the coefficients $\tau_{26}(p)$ can be computed in time polynomial in $\log(p)$ using a generalization of Cornacchia's algorithm, hence very fast.

Exercise: (You need some extra knowledge for this.) In the literature find an exact formula for $\tau_{26}(p)$ in terms of values of Hecke *Grössencharakters*, and program this formula. Use it to compute some values of $\tau_{26}(p)$ for p prime as large as you can go.

3.5. **Computing $|V(\mathbb{F}_q)|$ for Quasi-diagonal Hypersurfaces.** We now consider a completely different situation where $|V(\mathbb{F}_q)|$ can be computed without too much difficulty.

As we have seen, in the case of elliptic curves V defined over $\mathbb{Q}$, the corresponding L-function is of *degree* 2, in other words is of the form $\prod_p 1/(1-a(p)p^{-s}+b(p)p^{-2s})$, where $b(p)\neq 0$ for all but a finite number of p. L-functions of degree 1 such as the Riemann zeta function are essentially L-functions of Dirichlet characters, in other words simple "twists" of the Riemann zeta function. L-functions of degree 2 are believed to be always L-functions attached to modular forms, and $b(p)=\chi(p)p^{k-1}$ for a suitable integer k ($k=2$ for elliptic curves), the *weight*. Even though many unsolved questions remain, this case is also quite well understood. Much more mysterious are L-functions of higher degree, such as 3 or 4, and it is interesting to study natural mathematical objects leading to such functions. A case where this can be done reasonably easily is the case of diagonal or *quasi-diagonal hypersurfaces*. We study a special case:

Definition 3.2. *Let $m\ge 2$, for $1\le i\le m$ let $a_i\in\mathbb{F}_q^*$ be nonzero, and let $b\in\mathbb{F}_q$. The quasi-diagonal hypersurface defined by this data is the hypersurface in $\mathbb{P}^{m-1}$ defined by the projective equation*

$$\sum_{1\le i\le m} a_i x_i^m - b\prod_{1\le i\le m} x_i = 0\ .$$

When $b=0$, it is a diagonal hypersurface.

Of course, we could study more general equations, for instance where the degree is not equal to the number of variables, but we stick to this special case.

To compute the number of (projective) points on this hypersurface, we need an additional definition:

Definition 3.3. *We let ω be a generator of the group of characters of $\mathbb{F}_q^*$, either with values in $\mathbb{C}$, or in the p-adic field $\mathbb{C}_p$ (do not worry if you are not familiar with this).*

Indeed, by a well-known theorem of elementary algebra, the multiplicative group $\mathbb{F}_q^*$ of a finite field is *cyclic*, so its group of characters, which is *non-canonically isomorphic* to $\mathbb{F}_q^*$, is also cyclic, so ω indeed exists.

It is not difficult to prove the following theorem:

Theorem 3.4. *Assume that* $\gcd(m,q-1)=1$ *and $b\neq 0$, and set $B=\prod_{1\le i\le m}(a_i/b)$. If V is the above quasi-diagonal hypersurface, the number $|V(\mathbb{F}_q)|$ of* affine *points on V is given by*

$$|V(\mathbb{F}_q)| = q^{m-1}+(-1)^{m-1}+\sum_{1\le n\le q-2}\omega^{-n}(B)J_m(\omega^n,\dots,\omega^n)\ ,$$

where J_m is the m-variable Jacobi sum.

We will study in great detail below the definition and properties of J_m.

Note that the number of *projective* points is simply $(|V(\mathbb{F}_q)|-1)/(q-1)$.

There also exists a more general theorem with no restriction on $\gcd(m, q-1)$, which we do not give.

4. Gauss and Jacobi Sums

In this long section, we study in great detail Gauss and Jacobi sums. The emphasis will be on results which are not completely standard, the standard ones being stated without proof but with a reference. I would like to emphasize that almost all of these standard results can be proved with little difficulty by easy algebraic manipulations.

4.1. Gauss Sums over $\mathbb{F}_q$.

We can define and study Gauss and Jacobi sums in two different contexts: first, and most importantly, over finite fields $\mathbb{F}_q$, with $q = p^f$ a prime power (note that from now on we write $q = p^f$ and not $q = p^n$). Second, over the ring $\mathbb{Z}/N\mathbb{Z}$. The two notions coincide when $N = q = p$ is prime, but the methods and applications are quite different.

To give the definitions over $\mathbb{F}_q$ we need to recall some fundamental (and easy) results concerning finite fields.

Proposition 4.1. *Let p be a prime, $f \geq 1$, and $\mathbb{F}_q$ be the finite field with $q = p^f$ elements, which exists and is unique up to isomorphism.*

1. *The map ϕ such that $\phi(x) = x^p$ is a field isomorphism from $\mathbb{F}_q$ to itself leaving $\mathbb{F}_p$ fixed. It is called the* Frobenius map.
2. *The extension $\mathbb{F}_q/\mathbb{F}_p$ is a* normal *(i.e., separable and Galois) field extension, with Galois group which is cyclic of order f generated by ϕ.*

In particular, we can define the *trace* $\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}$ and the *norm* $\mathcal{N}_{\mathbb{F}_q/\mathbb{F}_p}$, and we have the formulas (where from now on we omit $\mathbb{F}_q/\mathbb{F}_p$ for simplicity):

$$\mathrm{Tr}(x) = \sum_{0 \leq j \leq f-1} x^{p^j} \quad \text{and} \quad \mathcal{N}(x) = \prod_{0 \leq j \leq f-1} x^{p^j} = x^{(p^f-1)/(p-1)} = x^{(q-1)/(p-1)} .$$

Definition 4.2. *Let χ be a character from $\mathbb{F}_q^*$ to an algebraically closed field C of characteristic 0. For $a \in \mathbb{F}_q$ we define the* Gauss sum $\mathfrak{g}(\chi, a)$ *by*

$$\mathfrak{g}(\chi, a) = \sum_{x \in \mathbb{F}_q^*} \chi(x) \zeta_p^{\mathrm{Tr}(ax)} ,$$

where ζ_p is a fixed primitive p-th root of unity in C. We also set $\mathfrak{g}(\chi) = \mathfrak{g}(\chi, 1)$.

Note that strictly speaking this definition depends on the choice of ζ_p. However, if ζ_p' is some other primitive p-th root of unity we have $\zeta_p' = \zeta_p^k$ for some $k \in \mathbb{F}_p^*$, so

$$\sum_{x \in \mathbb{F}_q^*} \chi(x) {\zeta_p'}^{\mathrm{Tr}(ax)} = \mathfrak{g}(\chi, ka) \ .$$

In fact it is trivial to see (this follows from the next proposition) that $\mathfrak{g}(\chi, ka) = \chi^{-1}(k)\mathfrak{g}(\chi, a)$.

Definition 4.3. *We define ε to be the trivial character, i.e., such that $\varepsilon(x) = 1$ for all $x \in \mathbb{F}_q^*$. We extend characters χ to the whole of $\mathbb{F}_q$ by setting $\chi(0) = 0$ if $\chi \neq \varepsilon$ and $\varepsilon(0) = 1$.*

Note that this apparently innocuous definition of $\varepsilon(0)$ is *crucial* because it simplifies many formulas. Note also that the definition of $\mathfrak{g}(\chi, a)$ is a sum over $x \in \mathbb{F}_q^*$ and not $x \in \mathbb{F}_q$, while for Jacobi sums we will use all of $\mathbb{F}_q$.

Exercise:

(1) Show that $\mathfrak{g}(\varepsilon, a) = -1$ if $a \in \mathbb{F}_q^*$ and $\mathfrak{g}(\varepsilon, 0) = q - 1$.

(2) If $\chi \neq \varepsilon$, show that $\mathfrak{g}(\chi, 0) = 0$, in other words that

$$\sum_{x \in \mathbb{F}_q} \chi(x) = 0$$

(here it does not matter if we sum over $\mathbb{F}_q$ or $\mathbb{F}_q^*$).

(3) Deduce that if $\chi_1 \neq \chi_2$ then

$$\sum_{x \in \mathbb{F}_q^*} \chi_1(x)\chi_2^{-1}(x) = 0 \ .$$

This relation is called for evident reasons *orthogonality of characters.*

Because of this exercise, if necessary we may assume that $\chi \neq \varepsilon$ and/or that $a \neq 0$.

Exercise: Let χ be a character of $\mathbb{F}_q^*$ of exact order n.

(1) Show that $n \mid (q-1)$ and that $\chi(-1) = (-1)^{(q-1)/n}$. In particular, if n is odd and $p > 2$ we have $\chi(-1) = 1$.

(2) Show that $\mathfrak{g}(\chi, a) \in \mathbb{Z}[\zeta_n, \zeta_p]$, where as usual ζ_m denotes a primitive mth root of unity.

Proposition 4.4. (1) *If $a \neq 0$ we have*

$$\mathfrak{g}(\chi, a) = \chi^{-1}(a)\mathfrak{g}(\chi) \ .$$

(2) *We have*

$$\mathfrak{g}(\chi^{-1}) = \chi(-1)\overline{\mathfrak{g}(\chi)} \ .$$

(3) *We have*

$$\mathfrak{g}(\chi^p, a) = \chi^{1-p}(a)\mathfrak{g}(\chi, a) \ .$$

(4) *If $\chi \neq \varepsilon$ we have*

$$|\mathfrak{g}(\chi)| = q^{1/2}\ .$$

4.2. **Jacobi Sums over $\mathbb{F}_q$.** Recall that we have extended characters of $\mathbb{F}_q^*$ by setting $\chi(0) = 0$ if $\chi \neq \varepsilon$ and $\varepsilon(0) = 1$.

Definition 4.5. *For $1 \leq j \leq k$ let χ_j be characters of $\mathbb{F}_q^*$. We define the Jacobi sum*

$$J_k(\chi_1, \dots, \chi_k; a) = \sum_{x_1+\cdots+x_k=a} \chi_1(x_1)\cdots\chi_k(x_k)$$

and $J_k(\chi_1, \dots, \chi_k) = J_k(\chi_1, \dots, \chi_k; 1)$.

Note that, as mentioned above, we do not exclude the cases where some $x_i = 0$, using the convention of Definition 4.3 for $\chi(0)$.

The following easy lemma shows that it is only necessary to study $J_k(\chi_1, \dots, \chi_k)$:

Lemma 4.6. *Set $\chi = \chi_1 \cdots \chi_k$.*

(1) *If $a \neq 0$ we have*

$$J_k(\chi_1, \dots, \chi_k; a) = \chi(a) J_k(\chi_1, \dots, \chi_k)\ .$$

(2) *If $a = 0$, abbreviating $J_k(\chi_1, \dots, \chi_k; 0)$ to $J_k(0)$ we have*

$$J_k(0) = \begin{cases} q^{k-1} & \text{if } \chi_j = \varepsilon \text{ for all } j\ , \\ 0 & \text{if } \chi \neq \varepsilon\ , \\ \chi_k(-1)(q-1)J_{k-1}(\chi_1, \dots, \chi_{k-1}) & \text{if } \chi = \varepsilon \text{ and } \chi_k \neq \varepsilon\ . \end{cases}$$

As we have seen, a Gauss sum $\mathfrak{g}(\chi)$ belongs to the rather large ring $\mathbb{Z}[\zeta_{q-1}, \zeta_p]$ (and in general not to a smaller ring). The advantage of Jacobi sums is that they belong to the smaller ring $\mathbb{Z}[\zeta_{q-1}]$, and as we are going to see, that they are closely related to Gauss sums. Thus, when working *algebraically*, it is almost always better to use Jacobi sums instead of Gauss sums. On the other hand, when working *analytically* (for instance in $\mathbb{C}$ or $\mathbb{C}_p$), it may be better to work with Gauss sums: we will see below the use of root numbers (suggested by Louboutin), and of the Gross–Koblitz formula.

Note that $J_1(\chi_1) = 1$. Outside of this trivial case, the close link between Gauss and Jacobi sums is given by the following easy proposition, whose apparently technical statement is only due to the trivial character ε: if none of the χ_j nor their product is trivial, we have the simple formula given by (3).

Proposition 4.7. *Denote by t the number of χ_j equal to the trivial character ε, and as above set $\chi = \chi_1 \dots \chi_k$.*

(1) *If $t = k$ then $J_k(\chi_1, \dots, \chi_k) = q^{k-1}$.*

(2) *If $1 \leq t \leq k-1$ then $J_k(\chi_1, \dots, \chi_k) = 0$.*

(3) *If $t = 0$ and $\chi \neq \varepsilon$ then*

$$J_k(\chi_1, \dots, \chi_k) = \frac{\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_k)}{\mathfrak{g}(\chi_1 \cdots \chi_k)} = \frac{\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_k)}{\mathfrak{g}(\chi)} \ .$$

(4) *If $t = 0$ and $\chi = \varepsilon$ then*

$$\begin{aligned} J_k(\chi_1, \dots, \chi_k) &= -\frac{\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_k)}{q} \\ &= -\chi_k(-1)\frac{\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_{k-1})}{\mathfrak{g}(\chi_1 \cdots \chi_{k-1})} = -\chi_k(-1)J_{k-1}(\chi_1, \dots, \chi_{k-1}) \ . \end{aligned}$$

In particular, in this case we have

$$\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_k) = \chi_k(-1)qJ_{k-1}(\chi_1, \dots, \chi_{k-1}) \ .$$

Corollary 4.8. *With the same notation, assume that $k \geq 2$ and all the χ_j are nontrivial. Setting $\psi = \chi_1 \cdots \chi_{k-1}$, we have the following recursive formula:*

$$J_k(\chi_1, \dots, \chi_k) = \begin{cases} J_{k-1}(\chi_1, \dots, \chi_{k-1})J_2(\psi, \chi_k) & \text{if } \psi \neq \varepsilon \ , \\ \chi_{k-1}(-1)qJ_{k-2}(\chi_1, \dots, \chi_{k-2}) & \text{if } \psi = \varepsilon \ . \end{cases}$$

The point of this recursion is that the definition of a k-fold Jacobi sum J_k involves a sum over q^{k-1} values for $x_1, \dots, x_{k-1}$, the last variable x_k being determined by $x_k = 1 - x_1 - \cdots - x_{k-1}$, so neglecting the time to compute the $\chi_j(x_j)$ and their product (which is a reasonable assumption), using the definition takes time $O(q^{k-1})$. On the other hand, using the above recursion boils down at worst to computing $k - 1$ Jacobi sums J_2, for a total time of $O((k-1)q)$. Nonetheless, we will see that in some cases it is still better to use directly Gauss sums and formula (3) of the proposition.

Since Jacobi sums J_2 are the simplest and the above recursion in fact shows that one can reduce to J_2, we will drop the subscript 2 and simply write $J(\chi_1, \chi_2)$. Note that

$$J(\chi_1, \chi_2) = \sum_{x \in \mathbb{F}_q} \chi_1(x)\chi_2(1-x) \ ,$$

where the sum is over the whole of $\mathbb{F}_q$ and *not* $\mathbb{F}_q \setminus \{0, 1\}$ (which makes a difference only if one of the χ_i is trivial). More precisely it is clear that $J(\varepsilon, \varepsilon) = q^2$, and that if $\chi \neq \varepsilon$ we have $J(\chi, \varepsilon) = \sum_{x \in \mathbb{F}_q} \chi(x) = 0$, which are special cases of Proposition 4.7.

Exercise: Let $n \mid (q-1)$ be the order of χ. Prove that $\mathfrak{g}(\chi)^n \in \mathbb{Z}[\zeta_n]$.

Exercise: Assume that none of the χ_j is equal to ε, but that their product χ is equal to ε. Prove that (using the same notation as in Lemma 4.6):

$$J_k(0) = \left(1 - \frac{1}{q}\right)\mathfrak{g}(\chi_1) \cdots \mathfrak{g}(\chi_k) \ .$$

Exercise: Prove the following reciprocity formula for Jacobi sums: if the χ_j are all nontrivial and $\chi = \chi_1 \cdots \chi_k$, we have

$$J_k(\chi_1^{-1}, \dots, \chi_k^{-1}) = \frac{q^{k-1-\delta}}{J_k(\chi_1, \dots, \chi_k)}\ ,$$

where $\delta = 1$ if $\chi = \varepsilon$, and otherwise $\delta = 0$.

4.3. **Applications of $J(\chi, \chi)$.** In this short subsection we give without proof a couple of applications of the special Jacobi sums $J(\chi, \chi)$. Once again the proofs are not difficult. We begin by the following result, which is a special case of the Hasse–Davenport relations that we will give below.

Lemma 4.9. *Assume that q is odd, and let ρ be the unique character of order 2 on $\mathbb{F}_q^*$. For any nontrivial character χ we have*

$$\chi(4)J(\chi, \chi) = J(\chi, \rho)\ .$$

Equivalently, if $\chi \neq \rho$ we have

$$\mathfrak{g}(\chi)\mathfrak{g}(\chi\rho) = \chi^{-1}(4)\mathfrak{g}(\rho)\mathfrak{g}(\chi^2)\ .$$

Exercise:

(1) Prove this lemma.
(2) Show that $\mathfrak{g}(\rho)^2 = (-1)^{(q-1)/2}q$.

Proposition 4.10. (1) *Assume that $q \equiv 1 \pmod 4$, let χ be one of the two characters of order 4 on $\mathbb{F}_q^*$, and write $J(\chi, \chi) = a + bi$. Then $q = a^2 + b^2$, $2 \mid b$, and $a \equiv -1 \pmod 4$.*
(2) *Assume that $q \equiv 1 \pmod 3$, let χ be one of the two characters of order 3 on $\mathbb{F}_q^*$, and write $J(\chi, \chi) = a + b\rho$, where $\rho = \zeta_3$ is a primitive cube root of unity. Then $q = a^2 - ab + b^2$, $3 \mid b$, $a \equiv -1 \pmod 3$, and $a + b \equiv q - 2 \pmod 9$.*
(3) *Let $p \equiv 2 \pmod 3$, $q = p^{2m} \equiv 1 \pmod 3$, and let χ be one of the two characters of order 3 on $\mathbb{F}_q^*$. We have*

$$J(\chi, \chi) = (-1)^{m-1}p^m = (-1)^{m-1}q^{1/2}\ .$$

Corollary 4.11. (1) *(Fermat.) Any prime $p \equiv 1 \pmod 4$ is a sum of two squares.*
(2) *Any prime $p \equiv 1 \pmod 3$ is of the form $a^2 - ab + b^2$ with $3 \mid b$, or equivalently $4p = (2a-b)^2 + 27(b/3)^2$ is of the form $c^2 + 27d^2$.*
(3) *(Gauss.) $p \equiv 1 \pmod 3$ is itself of the form $p = u^2 + 27v^2$ if and only if 2 is a cube in $\mathbb{F}_p^*$.*

Exercise: Assuming the proposition, prove the corollary.

4.4. **The Hasse–Davenport Relations.** All the results that we have given up to now on Gauss and Jacobi sums have rather simple proofs, which is one of the reasons we have not given them. Perhaps surprisingly, there exist other important relations which are considerably more difficult to prove. Before giving them, it is instructive to explain how one can "guess" their existence, if one knows the classical theory of the gamma function $\Gamma(s)$ (of course skip this part if you do not know it, since it would only confuse you).

Recall that $\Gamma(s)$ is defined (at least for $\Re(s) > 0$) by

$$\Gamma(s) = \int_0^\infty e^{-t} t^s dt/t \ ,$$

and the beta function $B(a,b)$ by $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$. The function e^{-t} transforms sums into products, so is an *additive* character, analogous to ζ_p^t. The function t^s transforms products into products, so is a multiplicative character, analogous to $\chi(t)$ (dt/t is simply the Haar invariant measure on $\mathbb{R}_{>0}$). Thus $\Gamma(s)$ is a continuous analogue of the Gauss sum $\mathfrak{g}(\chi)$.

Similarly, since $J(\chi_1,\chi_2) = \sum_t \chi_1(t)\chi_2(1-t)$, we see the similarity with the function B. Thus, it does not come too much as a surprise that analogous formulas are valid on both sides. To begin with, it is not difficult to show that $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, exactly analogous to $J(\chi_1,\chi_2) = \mathfrak{g}(\chi_1)\mathfrak{g}(\chi_2)/\mathfrak{g}(\chi_1\chi_2)$. The analogue of $\Gamma(s)\Gamma(-s) = -\pi/(s\sin(s\pi))$ is

$$\mathfrak{g}(\chi)\mathfrak{g}(\chi^{-1}) = \chi(-1)q \ .$$

But it is well-known that the gamma function has a duplication formula $\Gamma(s)\Gamma(s+1/2) = 2^{1-2s}\Gamma(1/2)\Gamma(2s)$, and more generally a multiplication (or distribution) formula. This duplication formula is clearly the analogue of the formula

$$\mathfrak{g}(\chi)\mathfrak{g}(\chi\rho) = \chi^{-1}(4)\mathfrak{g}(\rho)\mathfrak{g}(\chi^2)$$

given above. The *Hasse–Davenport product relation* is the analogue of the distribution formula for the gamma function.

Theorem 4.12. *Let ρ be a character of exact order m dividing $q-1$. For any character χ of $\mathbb{F}_q^*$ we have*

$$\prod_{0\le a<m} \mathfrak{g}(\chi\rho^a) = \chi^{-m}(m)k(p,f,m)q^{(m-1)/2}\mathfrak{g}(\chi^m) \ ,$$

where $k(p,f,m)$ is the fourth root of unity given by

$$k(p,f,m) = \begin{cases} \left(\dfrac{p}{m}\right)^f & \text{if } m \text{ is odd,} \\ (-1)^{f+1}\left(\dfrac{(-1)^{m/2+1}m/2}{p}\right)^f\left(\dfrac{-1}{p}\right)^{f/2} & \text{if } m \text{ is even,} \end{cases}$$

where $(-1)^{f/2}$ is to be understood as i^f when f is odd.

Remark: For some reason, in the literature this formula is usually stated in the weaker form where the constant $k(p,f,m)$ is not given explicitly.

Contrary to the proof of the distribution formula for the gamma function, the proof of this theorem is quite long. There are essentially two completely different proofs: one using classical algebraic number theory, and one using p-adic analysis. The latter is simpler and gives directly the value of $k(p,f,m)$. See [4] and [5] for both detailed proofs.

Gauss sums satisfy another type of nontrivial relation, also due to Hasse–Davenport, the so-called *lifting relation*, as follows:

Theorem 4.13. *Let $\mathbb{F}_{q^n}/\mathbb{F}_q$ be an extension of finite fields, let χ be a character of $\mathbb{F}_q^*$, and define the* lift *of χ to $\mathbb{F}_{q^n}$ by the formula $\chi^{(n)} = \chi \circ \mathcal{N}_{\mathbb{F}_{q^n}/\mathbb{F}_q}$. We have*

$$\mathfrak{g}(\chi^{(n)}) = (-1)^{n-1}\mathfrak{g}(\chi)^n \ .$$

This relation is essential in the initial proof of the Weil conjectures for diagonal hypersurfaces done by Weil himself. This is not surprising, since we have seen in Theorem 3.4 that $|V(\mathbb{F}_q)|$ is closely related to Jacobi sums, hence also to Gauss sums.

5. Practical Computations of Gauss and Jacobi Sums

As above, let ω be a character of order exactly $q-1$, so that ω is a generator of the group of characters of $\mathbb{F}_q^*$. For notational simplicity, we will write $J(r_1,\dots,r_k)$ instead of $J(\omega^{r_1},\dots,\omega^{r_k})$. Let us consider the specific example of efficient computation of the quantity

$$S(q;z) = \sum_{0\le n\le q-2} \omega^{-n}(z)J_5(n,n,n,n,n) \ ,$$

which occurs in the computation of the Hasse–Weil zeta function of a quasi-diagonal threefold, see Theorem 3.4.

5.1. Elementary Methods.

By the recursion of Corollary 4.8, we have *generically* (i.e., except for special values of n which will be considered separately):

$$J_5(n,n,n,n,n) = J(n,n)J(2n,n)J(3n,n)J(4n,n) \ .$$

Since $J(n,an) = \sum_x \omega^n(x)\omega^{an}(1-x)$, the cost of computing J_5 as written is $\widetilde{O}(q)$, where here and after we write $\widetilde{O}(q^\alpha)$ to mean $O(q^{\alpha+\varepsilon})$ for all $\varepsilon > 0$ (soft-O notation). Thus computing $S(q;z)$ by this direct method requires time $\widetilde{O}(q^2)$.

We can however do much better. Since the values of the characters are all in $\mathbb{Z}[\zeta_{q-1}]$, we work in this ring. In fact, even better, we work in the ring with zero divisors $R = \mathbb{Z}[X]/(X^{q-1}-1)$, together with the natural surjective map sending the class of X in R to ζ_{q-1}. Indeed, let g be the generator of $\mathbb{F}_q^*$ such that $\omega(g) = \zeta_{q-1}$. We have, again *generically*:

$$J(n, an) = \sum_{1\le u\le q-2} \omega^n(g^u)\omega^{an}(1-g^u) = \sum_{1\le u\le q-2} \zeta_{q-1}^{nu+an\log_g(1-g^u)}\ ,$$

where $\log_g$ is the *discrete logarithm* to base g defined modulo $q-1$, i.e., such that $g^{\log_g(x)} = x$. If $(q-1)\nmid n$ but $(q-1)\mid an$ we have $\omega^{an} = \varepsilon$ so we must add the contribution of $u = 0$, which is 1, and if $(q-1)\mid n$ we must add the contribution of $u = 0$ *and* of $x = 0$, which is 2 (recall the *essential* convention that $\chi(0) = 0$ if $\chi \neq \varepsilon$ and $\varepsilon(0) = 1$, see Definition 4.3).

In other words, if we set

$$P_a(X) = \sum_{1\le u\le q-2} X^{(u+a\log_g(1-g^u)) \bmod (q-1)} \in R\ ,$$

we have

$$J(n, an) = P_a(\zeta_{q-1}^n) + \begin{cases} 0 & \text{if } (q-1)\nmid an\ , \\ 1 & \text{if } (q-1)\mid an \text{ but } (q-1)\nmid n\ , \text{ and} \\ 2 & \text{if } (q-1)\mid n\ . \end{cases}$$

Thus, if we set finally

$$P(X) = P_1(X)P_2(X)P_3(X)P_4(X) \bmod X^{q-1} \in R\ ,$$

we have (still generically) $J_5(n, n, n, n, n) = P(\zeta_{q-1}^n)$. Assume for the moment that this is true for all n (we will correct this below), let $\ell = \log_g(z)$, so that $\omega(z) = \omega(g^\ell) = \zeta_{q-1}^\ell$, and write

$$P(X) = \sum_{0\le j\le q-2} a_j X^j\ .$$

We thus have

$$\omega^{-n}(z)J_5(n, n, n, n, n) = \zeta_{q-1}^{-n\ell} \sum_{0\le j\le q-2} a_j\zeta_{q-1}^{nj} = \sum_{0\le j\le q-2} a_j\zeta_{q-1}^{n(j-\ell)}\ ,$$

hence

$$\begin{aligned} S(q; z) = \sum_{0\le n\le q-2} \omega^{-n}(z)J_5(n, n, n, n, n) &= \sum_{0\le j\le q-2} a_j \sum_{0\le n\le q-2} \zeta_{q-1}^{n(j-\ell)} \\ &= (q-1) \sum_{0\le j\le q-2,\ j\equiv\ell \pmod{q-1}} a_j = (q-1)a_\ell\ . \end{aligned}$$

The result is thus immediate as soon as we know the coefficients of the polynomial P. Since there exist fast methods for computing discrete logarithms, this leads to a $\widetilde{O}(q)$ method for computing $S(q; z)$.

To obtain the correct formula, we need to adjust for the special n for which $J_5(n, n, n, n, n)$ is not equal to $J(n, n)J(n, 2n)J(n, 3n)J(n, 4n)$, which are the same for which $(q-1)\mid an$ for some a such that $2 \le a \le 4$, together with $a = 5$. This is easy but boring, and should be skipped on first reading.

(1) For $n = 0$ we have $J_5(n, n, n, n, n) = q^4$, and on the other hand $P(1) = (J(0, 0) - 2)^4 = (q - 2)^4$, so the correction term is $q^4 - (q - 2)^4 = 8(q - 1)(q^2 - 2q + 2)$.

(2) For $n = (q - 1)/2$ (if q is odd) we have

$$J_5(n, n, n, n, n) = \mathfrak{g}(\omega^n)^5/\mathfrak{g}(\omega^{5n}) = \mathfrak{g}(\omega^n)^4 = \mathfrak{g}(\rho)^4$$

since $5n \equiv n \pmod{q-1}$, where ρ is the character of order 2, and we have $\mathfrak{g}(\rho)^2 = (-1)^{(q-1)/2}q$, so $J_5(n, n, n, n, n) = q^2$. On the other hand

$$\begin{aligned}P(\zeta_{q-1}^n) &= J(\rho, \rho)(J(\rho, 2\rho) - 1)J(\rho, \rho)(J(\rho, 2\rho) - 1)\\ &= J(\rho, \rho)^2 = \mathfrak{g}(\rho)^4/q^2 = 1\ ,\end{aligned}$$

so the correction term is $\rho(z)(q^2 - 1)$.

(3) For $n = \pm(q - 1)/3$ (if $q \equiv 1 \pmod 3$), writing $\chi_3 = \omega^{(q-1)/3}$, which is one of the two cubic characters, we have

$$\begin{aligned}J_5(n, n, n, n, n) &= \mathfrak{g}(\omega^n)^5/\mathfrak{g}(\omega^{5n}) = \mathfrak{g}(\omega^n)^5/\mathfrak{g}(\omega^{-n})\\ &= \mathfrak{g}(\omega^n)^6/(\mathfrak{g}(\omega^{-n})\mathfrak{g}(\omega^n)) = \mathfrak{g}(\omega^n)^6/q\\ &= qJ(n, n)^2\end{aligned}$$

(check all this). On the other hand

$$\begin{aligned}P(\zeta_{q-1}^n) &= J(n, n)J(n, 2n)(J(n, 3n) - 1)J(n, 4n)\\ &= \frac{\mathfrak{g}(\omega^n)^2}{\mathfrak{g}(\omega^{2n})}\frac{\mathfrak{g}(\omega^n)\mathfrak{g}(\omega^{2n})}{q}\frac{\mathfrak{g}(\omega^n)^2}{\mathfrak{g}(\omega^{2n})}\\ &= \frac{\mathfrak{g}(\omega^n)^5}{q\mathfrak{g}(\omega^{-n})} = \frac{\mathfrak{g}(\omega^n)^6}{q^2} = J(n, n)^2\ ,\end{aligned}$$

so the correction term is $2(q - 1)\Re(\chi_3^{-1}(z)J(\chi_3, \chi_3)^2)$.

(4) For $n = \pm(q - 1)/4$ (if $q \equiv 1 \pmod 4$), writing $\chi_4 = \omega^{(q-1)/4}$, which is one of the two quartic characters, we have

$$J_5(n, n, n, n, n) = \mathfrak{g}(\omega^n)^5/\mathfrak{g}(\omega^{5n}) = \mathfrak{g}(\omega^n)^4 = \omega^n(-1)qJ_3(n, n, n)\ .$$

In addition, we have

$$J_3(n, n, n) = J(n, n)J(n, 2n) = \omega^n(4)J(n, n)^2 = \rho(2)J(n, n)^2\ ,$$

so

$$J_5(n, n, n, n, n) = \mathfrak{g}(\omega^n)^4 = \omega^n(-1)q\rho(2)J(n, n)^2\ .$$

Note that

$$\chi_4(-1) = \chi_4^{-1}(-1) = \rho(2) = (-1)^{(q-1)/4}\ ,$$

(Exercise: prove it!), so that $\omega^n(-1)\rho(2) = 1$ and the above simplifies to $J_5(n, n, n, n, n) = qJ(n, n)^2$.

On the other hand,

$$\begin{aligned}P(\zeta_{q-1}^n) &= J(n,n)J(n,2n)J(n,3n)(J(n,4n)-1)\\ &= \frac{\mathfrak{g}(\omega^n)^2}{\mathfrak{g}(\omega^{2n})}\frac{\mathfrak{g}(\omega^n)\mathfrak{g}(\omega^{2n})}{\mathfrak{g}(\omega^{3n})}\frac{\mathfrak{g}(\omega^n)\mathfrak{g}(\omega^{3n})}{q}\\ &= \frac{\mathfrak{g}(\omega^n)^4}{q} = \omega^n(-1)\rho(2)J(n,n)^2 = J(n,n)^2\end{aligned}$$

as above, so the correction term is $2(q-1)\Re(\chi_4^{-1}(z)J(\chi_4,\chi_4)^2)$.

(5) For $n = a(q-1)/5$ with $1 \le a \le 4$ (if $q \equiv 1 \pmod 5$), writing $\chi_5 = \omega^{(q-1)/5}$ we have $J_5(n,n,n,n,n) = -\mathfrak{g}(\chi_5^a)^5/q$, while abbreviating $\mathfrak{g}(\chi_5^{am})$ to $g(m)$ we have

$$\begin{aligned}P(\zeta_{q-1}^n) &= J(n,n)J(n,2n)J(n,3n)J(n,4n)\\ &= -\frac{g(n)^2}{g(2n)}\frac{g(n)g(2n)}{g(3n)}\frac{g(n)g(3n)}{g(4n)}\frac{g(n)g(4n)}{q}\\ &= -\frac{g(n)^5}{q}\ ,\end{aligned}$$

so there is no correction term.

Summarizing, we have shown the following:

Proposition 5.1. *Let* $S(q;z) = \sum_{0\le n\le q-2}\omega^{-n}(z)J_5(n,n,n,n,n)$. *Let* $\ell = \log_g(z)$ *and let* $P(X) = \sum_{0\le j\le q-2} a_jX^j$ *be the polynomial defined above. We have*

$$S(q;z) = (q-1)(T_1+T_2+T_3+T_4+a_\ell)\ ,$$

where $T_m = 0$ *if* $m \nmid (q-1)$ *and otherwise*

$$T_1 = 8(q^2-2q+2)\ ,\quad T_2 = \rho(z)(q+1)\ ,$$

$$T_3 = 2\Re(\chi_3^{-1}(z)J(\chi_3,\chi_3)^2)\ ,\quad \textit{and}\quad T_4 = 2\Re(\chi_4^{-1}(z)J(\chi_4,\chi_4)^2)\ ,$$

with the above notation.

Note that thanks to Proposition 4.10, these supplementary Jacobi sums $J(\chi_3,\chi_3)$ and $J(\chi_4,\chi_4)$ can be computed in logarithmic time using Cornacchia's algorithm (this is not quite true, one needs an additional slight computation, do you see why?).

Note also for future reference that the above proposition *proves* that $(q-1) \mid S(q,z)$, which is not clear from the definition.

5.2. **Sample Implementations.** For simplicity, assume that $q = p$ is prime. I have written simple implementations of the computation of $S(q;z)$. In the first implementation, I use the naïve formula expressing J_5 in terms of $J(n,an)$ and sum on n, except that I use the reciprocity formula which gives $J_5(-n,-n,-n,-n,-n)$ in terms of $J_5(n,n,n,n,n)$ to sum only over $(p-1)/2$ terms instead of $p-1$. Of course to avoid recomputation, I precompute a discrete logarithm table.

The timings for $p \approx 10^k$ for $k = 2$, 3, and 4 are 0.03, 1.56, and 149 seconds respectively, compatible with $\widetilde{O}(q^2)$ time.

On the other hand, implementing in a straightforward manner the algorithm given by the above proposition gives timings for $p \approx 10^k$ for $k = 2$, 3, 4, 5, 6, and 7 of 0, 0.02, 0.08, 0.85, 9.90, and 123 seconds respectively, of course much faster and compatible with $\widetilde{O}(q)$ time.

The main drawback of this method is that it requires $O(q)$ storage: it is thus applicable only for $q \le 10^8$, say, which is more than sufficient for many applications, but of course not for all. For instance, the case $p \approx 10^7$ mentioned above already required a few gigabytes of storage.

5.3. **Using Theta Functions.** A completely different way of computing Gauss and Jacobi sums has been suggested by Louboutin. It is related to the theory of L-functions of Dirichlet characters that we study below, and in our context is valid only for $q = p$ prime, not for prime powers, but in the context of Dirichlet characters it is valid in general (simply replace p by N and $\mathbb{F}_p$ by $\mathbb{Z}/N\mathbb{Z}$ in the following formulas when χ is a primitive character of conductor N, see below for definitions):

Definition 5.2. *Let χ be a character on $\mathbb{F}_p$, and let $e = 0$ or 1 be such that $\chi(-1) = (-1)^e$. The* theta function *associated to χ is the function defined on the upper half-plane by*

$$\Theta(\chi, \tau) = 2\sum_{m\ge1} m^e \chi(m) e^{i\pi m^2\tau/p}\ .$$

The main property of this function, which is a direct consequence of the *Poisson summation formula*, and is equivalent to the functional equation of Dirichlet L-functions, is as follows:

Proposition 5.3. *We have the functional equation*

$$\Theta(\chi, -1/\tau) = W(\chi)(\tau/i)^{(2e+1)/2}\Theta(\chi^{-1}, \tau)\ ,$$

with the principal determination of the square root, and where $W(\chi) = \mathfrak{g}(\chi)/(i^e p^{1/2})$ is the so-called root number.

Corollary 5.4. *If $\chi(-1) = 1$ we have*

$$\mathfrak{g}(\chi) = p^{1/2}\frac{\sum_{m\ge1}\chi(m)\exp(-\pi m^2/pt)}{t^{1/2}\sum_{m\ge1}\chi^{-1}(m)\exp(-\pi m^2 t/p)}$$

and if $\chi(-1) = -1$ we have

$$\mathfrak{g}(\chi) = p^{1/2}i\frac{\sum_{m\ge1}\chi(m)m\exp(-\pi m^2/pt)}{t^{3/2}\sum_{m\ge1}\chi^{-1}(m)m\exp(-\pi n^2 t/p)}$$

for any t such that the denominator does not vanish.

Note that the optimal choice of t is $t = 1$, and (at least for p prime) it seems that the denominator never vanishes (there are counterexamples when p is not prime, but apparently only four, see [7]).

It follows from this corollary that $\mathfrak{g}(\chi)$ can be computed numerically as a complex number in $\widetilde{O}(p^{1/2})$ operations. Thus, if χ_1 and χ_2 are nontrivial characters such that $\chi_1\chi_2 \neq \varepsilon$ (otherwise $J(\chi_1, \chi_2)$ is trivial to compute), the formula $J(\chi_1, \chi_2) = \mathfrak{g}(\chi_1)\mathfrak{g}(\chi_2)/\mathfrak{g}(\chi_1\chi_2)$ allows the computation of J_2 *numerically* as a complex number in $\widetilde{O}(p^{1/2})$ operations.

To recover J itself as an algebraic number we could either compute all its conjugates, but this would require more time than the direct computation of J, or possibly use the LLL algorithm, which although fast, would also require some time. In practice, if we proceed as above, we only need J to sufficient accuracy: we perform all the elementary operations in $\mathbb{C}$, and since we know that at the end the result will be an integer for which we know an upper bound, we thus obtain a proven exact result.

More generally, we have generically $J_5(n, n, n, n, n) = \mathfrak{g}(\omega^n)^5/\mathfrak{g}(\omega^{5n})$, which can thus be computed in $\widetilde{O}(p^{1/2})$ operations. It follows that $S(p; z)$ can be computed in $\widetilde{O}(p^{3/2})$ operations, which is slower than the elementary method seen above. The main advantage is that we do not need much storage: more precisely, we want to compute $S(p; z)$ to sufficiently small accuracy that we can recognize it as an integer, so a priori up to an absolute error of 0.5. However, we have seen that $(p-1) \mid S(p; z)$: it is thus sufficient to have an absolute error less than $(p-1)/2$ thus at worse each of the $p-1$ terms in the sum to an absolute error less than $1/2$. Since generically $|J_5(n, n, n, n, n)| = p^2$, we need a relative error less than $1/(2p^2)$, so less than $1/(10p^2)$ on each Gauss sum. In practice of course this is overly pessimistic, but it does not matter. For $p \leq 10^9$, this means that 19 decimal digits suffice.

The main term in the theta function computation (with $t = 1$) is $\exp(-\pi m^2/p)$, so we need $\exp(-\pi m^2/p) \leq 1/(100p^2)$, say, in other words $\pi m^2/p \geq 4.7 + 2\log(p)$, so $m^2 \geq p(1.5 + 0.7\log(p))$.

This means that we will need the values of $\omega(m)$ only up to this limit, of the order of $O((p\log(p))^{1/2})$, considerably smaller than p. Thus, instead of computing a full discrete logarithm table, which takes some time but more importantly a lot of space, we compute only discrete logarithms up to that limit, using specific algorithms for doing so which exist in the literature, some of which being quite easy.

A straightforward implementation of this method gives timings for $k = 2$, 3, 4, and 5 of 0.02, 0.40, 16.2, and 663 seconds respectively, compatible with $\widetilde{O}(p^{3/2})$ time. This is faster than the completely naïve method, but slower than the method explained above. Its advantage is that it requires much less space. For p around 10^7, however, it is much too slow so this method is not much use. We will see that its usefulness is mainly in the context where it was invented, i.e., for L-functions of Dirichlet characters.

5.4. **Using the Gross–Koblitz Formula.** This section is of a considerably higher mathematical level than the preceding ones, but is very important since it gives the best method for computing Gauss (and Jacobi) sums. We refer to [5] for complete details, and urge the reader to try to understand what follows.

In the preceding sections, we have considered Gauss sums as belonging to a number of different rings: the ring $\mathbb{Z}[\zeta_{q-1}, \zeta_p]$ or the field $\mathbb{C}$ of complex numbers, and for Jacobi sums the ring $\mathbb{Z}[\zeta_{q-1}]$, but also the ring $\mathbb{Z}[X]/(X^{q-1} - 1)$, and again the field $\mathbb{C}$.

In number theory there exist other algebraically closed fields which are useful in many contexts, the fields $\mathbb{C}_\ell$ of ℓ-adic numbers, one for each prime number ℓ. These fields come with a topology and analysis which are rather special: one of the main things to remember is that a sequence of elements tends to 0 if and only the ℓ-adic valuation of the elements (the largest exponent of ℓ dividing them) tends to infinity. For instance 2^m tends to 0 in $\mathbb{C}_2$, but in no other $\mathbb{C}_\ell$, and 15^m tends to 0 in $\mathbb{C}_3$ and in $\mathbb{C}_5$.

The most important subrings of $\mathbb{C}_\ell$ are the ring $\mathbb{Z}_\ell$ of ℓ-adic integers, the elements of which can be written as $x = a_0 + a_1\ell + \cdots + a_k\ell^k + \cdots$ with $a_j \in [0, \ell - 1]$, and its field of fractions $\mathbb{Q}_\ell$, which contains $\mathbb{Q}$.

In dealing with Gauss and Jacobi sums over $\mathbb{F}_q$ with $q = p^f$, the only $\mathbb{C}_\ell$ which is of use for us is the one with $\ell = p$ (in highbrow language, we are going to use implicitly *crystalline* p-adic methods, while for $\ell \neq p$ it would be *étale* ℓ-adic methods).

Apart from this relatively strange topology, many definitions and results valid on $\mathbb{C}$ have analogues in $\mathbb{C}_p$. The main object that we will need in our context is the analogue of the gamma function, naturally called the p-adic gamma function, in the present case due to Morita (there is another one, see [5]), and denoted Γ_p. Its definition is in fact quite simple:

Definition 5.5. *For $s \in \mathbb{Z}_p$ we define*

$$\Gamma_p(s) = \lim_{m \to s} (-1)^m \prod_{\substack{0 \le k < m \\ p \nmid k}} k \ ,$$

where the limit is taken over any sequence of positive integers m tending to s for the p-adic topology.

It is of course necessary to show that this definition makes sense, but this is not difficult, and most of the important properties of $\Gamma_p(s)$, analogous to those of $\Gamma(s)$, can be deduced from it.

However we need a much deeper property of $\Gamma_p(s)$ known as the Gross–Koblitz formula: it is in fact an analogue of a formula for $\Gamma(s)$ known as the Chowla–Selberg formula, and it is also closely related to the Davenport–Hasse relations that we have seen above.

The proof of the GK formula was initially given using tools of crystalline cohomology, but an elementary proof due to Robert now exists, see for instance [5] once again.

The GK formula tells us that certain products of p-adic gamma functions at *rational* arguments are in fact *algebraic numbers*, more precisely *Gauss sums* (explaining their importance for us). This is quite surprising since usually transcendental functions such as Γ_p take transcendental values.

To give a specific example, we have $\Gamma_5(1/4)^2 = -2 + \sqrt{-1}$, where $\sqrt{-1}$ is the square root in $\mathbb{Z}_5$ congruent to 3 modulo 5.

Before stating the formula we need to collect a number of facts, both on classical algebraic number theory and on p-adic analysis. None are difficult to prove, see [4] and [5]. Recall that $q = p^f$.

- We let $K = \mathbb{Q}(\zeta_p)$ and $L = K(\zeta_{q-1}) = \mathbb{Q}(\zeta_{q-1}, \zeta_p) = \mathbb{Q}(\zeta_{p(q-1)})$, so that L/K is an extension of degree $\phi(q-1)$. There exists a unique prime ideal $\mathfrak{p}$ of K above p, and we have $\mathfrak{p} = (1-\zeta_p)\mathbb{Z}_K$ and $\mathfrak{p}^{p-1} = p\mathbb{Z}_K$, and $\mathbb{Z}_K/\mathfrak{p} \simeq \mathbb{F}_p$. The prime ideal $\mathfrak{p}$ splits into a product of $g = \phi(q-1)/f$ prime ideals $\mathfrak{P}_j$ of degree f in the extension L/K, i.e., $\mathfrak{p}\mathbb{Z}_L = \mathfrak{P}_1 \cdots \mathfrak{P}_g$, and for any prime ideal $\mathfrak{P} = \mathfrak{P}_j$ we have $\mathbb{Z}_L/\mathfrak{P} \simeq \mathbb{F}_q$.

Exercise: Prove directly that for any f we have $f \mid \phi(p^f - 1)$.

- Fix one of the prime ideals $\mathfrak{P}$ as above. There exists a unique group isomorphism $\omega = \omega_{\mathfrak{P}}$ from $(\mathbb{Z}_L/\mathfrak{P})^*$ to the group of $(q-1)$st roots of unity in L, such that for all $x \in (\mathbb{Z}_L/\mathfrak{P})^*$ we have $\omega(x) \equiv x \pmod{\mathfrak{P}}$. It is called the *Teichmüller character*, and it can be considered as a character of order $q-1$ on $\mathbb{F}_q^* \simeq (\mathbb{Z}_L/\mathfrak{P})^*$. We can thus *instantiate* the definition of a Gauss sum over $\mathbb{F}_q$ by defining it as $\mathfrak{g}(\omega_{\mathfrak{P}}^{-r}) \in L$.

- Let ζ_p be a primitive pth root of unity in $\mathbb{C}_p$, fixed once and for all. There exists a unique $\pi \in \mathbb{Z}[\zeta_p]$ satisfying $\pi^{p-1} = -p$, $\pi \equiv 1 - \zeta_p \pmod{\pi^2}$, and we set $K_{\mathfrak{p}} = \mathbb{Q}_p(\pi) = \mathbb{Q}_p(\zeta_p)$, and $L_{\mathfrak{P}}$ the *completion* of L at $\mathfrak{P}$. The field extension $L_{\mathfrak{P}}/K_{\mathfrak{p}}$ is Galois, with Galois group isomorphic to $\mathbb{Z}/f\mathbb{Z}$ (which is the same as the Galois group of $\mathbb{F}_q/\mathbb{F}_p$, where $\mathbb{F}_p$ (resp., $\mathbb{F}_q$) is the so-called *residue field* of K (resp., L)).

- We set the following:

Definition 5.6. *We define the* p-adic Gauss sum *by*

$$\mathfrak{g}_q(r) = \sum_{x \in L_{\mathfrak{P}},\ x^{q-1}=1} x^{-r} \zeta_p^{\mathrm{Tr}_{L_{\mathfrak{P}}/K_{\mathfrak{p}}}(x)} \in L_{\mathfrak{P}}\ .$$

Note that this depends on the choice of ζ_p, or equivalently of π. Since $\mathfrak{g}_q(r)$ and $\mathfrak{g}(\omega_{\mathfrak{P}}^{-r})$ are algebraic numbers, it is clear that they are equal, although viewed in fields having different topologies. Thus, results about $\mathfrak{g}_q(r)$ translate immediately into results about $\mathfrak{g}(\omega_{\mathfrak{P}}^{-r})$, hence about general Gauss sums over finite fields.

The Gross–Koblitz formula is as follows:

Theorem 5.7 (Gross–Koblitz)**.** *Denote by $s(r)$ the sum of digits to base p of the integer $r \bmod (q-1)$, i.e., of the unique integer r' such that $r' \equiv r \pmod{q-1}$ and $0 \le r' < q-1$. We have*

$$\mathfrak{g}_q(r) = -\pi^{s(r)} \prod_{0 \le i < f} \Gamma_p\left(\left\{\frac{p^{f-i}r}{q-1}\right\}\right) ,$$

where $\{x\}$ denotes the fractional part of x.

Let us show how this can be used to compute Gauss or Jacobi sums, and in particular our sum $S(q;z)$. Assume for simplicity that $f = 1$, in other words that $q = p$: the right hand side is thus equal to $-\pi^{s(r)}\Gamma_p(\{pr/(p-1)\})$. Since we can always choose r such that $0 \le r < p-1$, we have $s(r) = r$ and $\{pr/(p-1)\} = \{r + r/(p-1)\} = r/(p-1)$, so the RHS is $-\pi^r\Gamma_p(r/(p-1))$. Now an easy property of Γ_p is that it is differentiable: recall that p is "small" in the p-adic topology, so $r/(p-1)$ is close to $-r$, more precisely $r/(p-1) = -r + pr/(p-1)$ (this is how we obtained it in the first place!). Thus in particular, if $p > 2$ we have the Taylor expansion

$$\begin{aligned}\Gamma_p(r/(p-1)) &= \Gamma_p(-r) + (pr/(p-1))\Gamma_p'(-r) + O(p^2)\\ &= \Gamma_p(-r) - pr\Gamma_p'(-r) + O(p^2) .\end{aligned}$$

Since $\mathfrak{g}_q(r)$ depends only on r modulo $p-1$, we will assume that $0 \le r < p-1$. In that case it is easy to show from the definition that

$$\Gamma_p(-r) = 1/r! \quad \text{and} \quad \Gamma_p'(-r) = (-\gamma_p + H_r)/r! ,$$

where $H_r = \sum_{1 \le n \le r} 1/n$ is the harmonic sum, and $\gamma_p = -\Gamma_p'(0)$ is the p-adic analogue of Euler's constant.

Exercise: Prove these formulas, as well as the congruence for γ_p given below.

There exist infinite (p-adic) series enabling accurate computation of γ_p, but since we only need it modulo p, we use the easily proved congruence $\gamma_p \equiv ((p-1)!+1)/p = W_p \pmod{p}$, the so-called *Wilson quotient.*

Thus the GK formula tells us that for $0 \le r < p-1$ we have

$$\mathfrak{g}_q(r) = -\frac{\pi^r}{r!}(1 - pr(H_r - W_p) + O(p^2)) .$$

It follows that for $(p-1) \nmid 5r$ we have

$$J(-r,-r,-r,-r,-r) = \frac{\mathfrak{g}(\omega_{\mathfrak{P}})^5}{\mathfrak{g}(\omega_{\mathfrak{P}}^5)} = \frac{\mathfrak{g}_q(r)^5}{\mathfrak{g}_q(5r)} = \pi^{f(r)}(a + bp + O(p^2)) ,$$

where a and b will be computed below and

$$\begin{aligned} f(r) &= 5r - (5r \bmod p-1) = 5r - (5r - (p-1)\lfloor 5r/(p-1)\rfloor) \\ &= (p-1)\lfloor 5r/(p-1)\rfloor\ , \end{aligned}$$

so that $\pi^{f(r)} = (-p)^{\lfloor 5r/(p-1)\rfloor}$ since $\pi^{p-1} = -p$. Since we want the result modulo p^2, we consider three intervals together with special cases:

(1) If $r > 2(p-1)/5$ but $(p-1) \nmid 5r$, we have

$$J(-r,-r,-r,-r,-r) \equiv 0 \pmod{p^2}\ .$$

(2) If $(p-1)/5 < r < 2(p-1)/5$ we have

$$J(-r,-r,-r,-r,-r) \equiv (-p)\frac{(5r-(p-1))!}{r!^5} \pmod{p^2}\ .$$

(3) If $0 < r < (p-1)/5$ we have $f(r) = 0$ and $0 \le 5r < (p-1)$ hence

$$\begin{aligned} J(-r,-r,-r,-r,-r) &= \frac{(5r)!}{r!^5}(1 - 5pr(H_r - W_p) + O(p^2))\cdot \\ &\quad \cdot (1 + 5pr(H_{5r} - W_p) + O(p^2)) \\ &\equiv \frac{(5r)!}{r!^5}(1 + 5pr(H_{5r} - H_r)) \pmod{p^2}\ . \end{aligned}$$

(4) Finally, if $r = a(p-1)/5$ we have $J(-r,-r,-r,-r,-r) = p^4 \equiv 0 \pmod{p^2}$ if $a = 0$, and otherwise $J(-r,-r,-r,-r,-r) = -\mathfrak{g}_q(r)^5/p$, and since the p-adic valuation of $\mathfrak{g}_q(r)$ is equal to $r/(p-1) = a/5$, that of $J(-r,-r,-r,-r,-r)$ is equal to $a-1$, which is greater or equal to 2 as soon as $a \ge 3$. For $a = 2$, i.e., $r = 2(p-1)/5$, we thus have

$$J(-r,-r,-r,-r,-r) \equiv p\frac{1}{r!^5} \equiv (-p)\frac{(5r-(p-1))!}{r!^5} \pmod{p^2}\ ,$$

which is the same formula as for $(p-1)/5 < r \le 2(p-1)/5$. For $a = 1$, i.e., $r = (p-1)/5$, we thus have

$$J(-r,-r,-r,-r,-r) \equiv -\frac{1}{r!^5}(1 - 5pr(H_r - W_p)) \pmod{p^2}\ ,$$

while on the other hand

$$(5r)! = (p-1)! = -1 + pW_p \equiv -1 - p(p-1)W_p \equiv -1 - 5prW_p\ ,$$

and $H_{5r} = H_{p-1} \equiv 0 \pmod{p}$ (Wolstenholme's congruence, easy), so

$$\begin{aligned} \frac{(5r)!}{r!^5}(1 + 5pr(H_{5r} - H_r)) &\equiv -\frac{1}{r!^5}(1 - 5prH_r)(1 + 5prW_p) \\ &\equiv -\frac{1}{r!^5}(1 - 5pr(H_r - W_p)) \pmod{p^2}\ , \end{aligned}$$

which is the same formula as for $0 < r < (p-1)/5$.

An important point to note is that we are working p-adically, but the final result $S(p;z)$ being an integer, it does not matter at the end. There is one small additional detail to take care of: we have

$$\begin{aligned} S(p;z) &= \sum_{0\le r\le p-2} \omega^{-r}(z)J(r,r,r,r,r) \\ &= \sum_{0\le r\le p-2} \omega^{r}(z)J(-r,-r,-r,-r,-r)\ , \end{aligned}$$

so we must express $\omega^r(z)$ in the p-adic setting. Since $\omega = \omega_{\mathfrak{P}}$ is the *Teichmüller character*, in the p-adic setting it is easy to show that $\omega(z)$ is the p-adic limit of z^{p^k} as $k \to \infty$. in particular $\omega(z) \equiv z \pmod{p}$, but more precisely $\omega(z) \equiv z^p \pmod{p^2}$.

Exercise: Let $p \ge 3$. Assume that $z \in \mathbb{Z}_p \setminus p\mathbb{Z}_p$ (for instance that $z \in \mathbb{Z} \setminus p\mathbb{Z}$). Prove that z^{p^k} has a p-adic limit $\omega(z)$ when $k \to \infty$, that $\omega^{p-1}(z) = 1$, that $\omega(z) \equiv z \pmod{p}$, and $\omega(z) \equiv z^p \pmod{p^2}$.

We have thus proved the following

Proposition 5.8. *We have*

$$\begin{aligned} S(p;z) \equiv &\sum_{0<r\le (p-1)/5} \frac{(5r)!}{r!^5}(1+5pr(H_{5r}-H_r))z^{pr} \\ &- p \sum_{(p-1)/5<r\le 2(p-1)/5} \frac{(5r-(p-1))!}{r!^5} z^r \pmod{p^2}\ . \end{aligned}$$

In particular

$$S(p;z) \equiv \sum_{0<r\le (p-1)/5} \frac{(5r)!}{r!^5} z^r \pmod{p}\ .$$

Remarks.

(1) Note that, as must be the case, all mention of p-adic numbers has disappeared from this formula. We used the p-adic setting only in the proof. It can be proved "directly", but with some difficulty.

(2) We used the Taylor expansion only to order 2. It is of course possible to use it to any order, thus giving a generalization of the above proposition to any power of p.

The point of giving all these details is as follows: it is easy to show that $(p-1) \mid S(p;z)$ (in fact we have seen this in the elementary method above). We can thus easily compute $S(p;z)$ modulo $p^2(p-1)$. On the other hand, it is possible to prove (but not easy, it is part of the Weil conjectures proved by Deligne), that $|S(p;z) - p^4| < 4p^{5/2}$. It follows that as soon as $8p^{5/2} < p^2(p-1)$, in other words $p \ge 67$, the computation that we perform modulo p^2 is sufficient to determine

$S(p;z)$ exactly. It is clear that the time to perform this computation is $\widetilde{O}(p)$, and in fact much faster than any that we have seen.

In fact, implementing in a reasonable way the algorithm given by the above proposition gives timings for $p \approx 10^k$ for $k = 2$, 3, 4, 5, 6, 7, and 8 of 0, 0.01, 0.03, 0.21, 2.13, 21.92, and 229.6 seconds respectively, of course much faster and compatible with $\widetilde{O}(p)$ time. The great additional advantage is that we use very small memory. This is therefore the best known method.

Numerical example: Choose $p = 10^6 + 3$ and $z = 2$. In 2.13 seconds we find that $S(p;z) \equiv a \pmod{p^2}$ with $a = 356022712041$. Using the Chinese remainder formula

$$S(p;z) = p^4 + ((a - (1+a)p^2) \bmod ((p-1)p^2))\ ,$$

we immediately deduce that

$$S(p;z) = 1000012000056356142712140\ .$$

Here is a summary of the timings (in seconds) that we have mentioned:

k	2	3	4	5	6	7	8
Naïve	0.03	1.56	149	$*$	$*$	$*$	$*$
Theta	0.02	0.40	16.2	663	$*$	$*$	$*$
Mod $X^{q-1}-1$	0	0.02	0.08	0.85	9.90	123	$*$
Gross–Koblitz	0	0.01	0.03	0.21	2.13	21.92	229.6

Time for computing $S(p;z)$ for $p \approx 10^k$

6. Gauss and Jacobi Sums over $\mathbb{Z}/N\mathbb{Z}$

Another context in which one encounters Gauss sums is over finite rings such as $\mathbb{Z}/N\mathbb{Z}$. The theory coincides with that over $\mathbb{F}_q$ when $q = p = N$ is prime, but is rather different otherwise. These other Gauss sums enter in the important theory of *Dirichlet characters.*

6.1. **Definitions.** We recall the following definition:

Definition 6.1. *Let χ be a (multiplicative) character from the multiplicative group $(\mathbb{Z}/N\mathbb{Z})^*$ of invertible elements of $\mathbb{Z}/N\mathbb{Z}$ to the complex numbers $\mathbb{C}$. We denote by abuse of notation again by χ the map from $\mathbb{Z}$ to $\mathbb{C}$ defined by $\chi(x) = \chi(x \bmod N)$ when x is coprime to N, and $\chi(x) = 0$ if x is not coprime to N, and call it the Dirichlet character modulo N associated to χ.*

It is clear that a Dirichlet character satisfies $\chi(xy) = \chi(x)\chi(y)$ for all x and y, that $\chi(x+N) = \chi(x)$, and that $\chi(x) = 0$ if and only if x is not coprime with N. Conversely, it immediate that these properties characterize Dirichlet characters.

A crucial notion (which has no equivalent in the context of characters of $\mathbb{F}_q^*$) is that of *primitivity*:

Assume that $M \mid N$. If χ is a Dirichlet character modulo M, we can transform it into a character χ_N modulo N by setting $\chi_N(x) = \chi(x)$ if x is coprime to N, and $\chi_N(x) = 0$ otherwise. We say that the characters χ and χ_N are *equivalent*. Conversely, if ψ is a character modulo N, it is not always true that one can find χ modulo M such that $\psi = \chi_N$. If it is possible, we say that ψ *can be defined modulo* M.

Definition 6.2. *Let χ be a character modulo N. We say that χ is a* primitive character *if χ cannot be defined modulo M for any proper divisor M of N, i.e., for any $M \mid N$ such that $M \neq N$.*

Exercise: Assume that $N \equiv 2 \pmod 4$. Show that there do not exist any primitive characters modulo N.

Exercise: Assume that $p^a \mid N$. Show that if χ is a primitive character modulo N, the *order* of χ (the smallest k such that χ^k is a trivial character) is *divisible* by p^{a-1}.

As we will see, questions about general Dirichlet characters can always be reduced to questions about primitive characters, and the latter have much nicer properties.

Proposition 6.3. *Let χ be a character modulo N. There exists a divisor f of N called the* conductor *of χ (this f has nothing to do with the f used above such that $q = p^f$), having the following properties:*

1. *The character χ can be defined modulo f, in other words there exists a character ψ modulo f such that $\chi = \psi_N$ using the notation above.*
2. *f is the smallest divisor of N having this property.*
3. *The character ψ is a primitive character modulo f.*

There is also the notion of *trivial character modulo N*: however we must be careful here, and we set the following:

Definition 6.4. *The trivial character modulo N is the Dirichlet character associated with the trivial character of $(\mathbb{Z}/N\mathbb{Z})^*$. It is usually denoted by χ_0 (but be careful, the index N is implicit, so χ_0 may represent different characters), and its values are as follows: $\chi_0(x) = 1$ if x is coprime to N, and $\chi_0(x) = 0$ if x is not coprime to N.*

In particular, $\chi_0(0) = 0$ if $N \neq 1$. The character χ_0 can also be characterized as the only character modulo N of conductor 1.

Definition 6.5. *Let χ be a character modulo N. The* Gauss sum *associated to χ and $a \in \mathbb{Z}$ is*

$$\mathfrak{g}(\chi, a) = \sum_{x \bmod N} \chi(x)\zeta_N^{ax}\ ,$$

and we write simply $\mathfrak{g}(\chi)$ instead of $\mathfrak{g}(\chi, 1)$.

The most important results concerning these Gauss sums is the following:

Proposition 6.6. *Let χ be a character modulo N.*

1. *If a is coprime to N we have*
$$\mathfrak{g}(\chi,a)=\chi^{-1}(a)\mathfrak{g}(\chi)=\overline{\chi(a)}\mathfrak{g}(\chi)\ ,$$
and more generally $\mathfrak{g}(\chi,ab)=\chi^{-1}(a)\mathfrak{g}(\chi,b)=\overline{\chi(a)}\mathfrak{g}(\chi,b)$.
2. *If χ is a* primitive *character, we have*
$$\mathfrak{g}(\chi,a)=\overline{\chi(a)}\mathfrak{g}(\chi)$$
for all *a, in other words, in addition to (1), we have $\mathfrak{g}(\chi,a)=0$ if a is not coprime to N.*
3. *If χ is a* primitive *character, we have $|\mathfrak{g}(\chi)|^2=N$.*

Note that (1) is trivial, and that since $\chi(a)$ has modulus 1 when a is coprime to N, we can write indifferently $\chi^{-1}(a)$ or $\overline{\chi(a)}$. On the other hand, (2) is not completely trivial.

We leave to the reader the easy task of defining Jacobi sums and of proving the easy relations between Gauss and Jacobi sums.

6.2. **Reduction to Prime Gauss Sums.**

A fundamental and little-known fact is that in the context of Gauss sums over $\mathbb{Z}/N\mathbb{Z}$ (as opposed to $\mathbb{F}_q$), one can in fact always reduce to prime N. First note (with proof) the following easy result:

Proposition 6.7. *Let $N=N_1N_2$ with N_1 and N_2 coprime, and let χ be a character modulo N.*

1. *There exist unique characters χ_i modulo N_i such that $\chi=\chi_1\chi_2$ in an evident sense, and if χ is primitive, the χ_i will also be primitive.*
2. *We have the identity (valid even if χ is not primitive):*
$$\mathfrak{g}(\chi)=\chi_1(N_2)\chi_2(N_1)\mathfrak{g}(\chi_1)\mathfrak{g}(\chi_2)\ .$$

Proof. (1). Since N_1 and N_2 are coprime there exist u_1 and u_2 such that $u_1N_1+u_2N_2=1$. We define $\chi_1(x)=\chi(xu_2N_2+u_1N_1)$ and $\chi_2(x)=\chi(xu_1N_1+u_2N_2)$. We leave to the reader to check (1) using these definitions.

(2). When x_i ranges modulo N_i, $x=x_1u_2N_2+x_2u_1N_1$ ranges modulo N (check it, in particular that the values are distinct!), and $\chi(x)=\chi_1(x)\chi_2(x)=\chi_1(x_1)\chi_2(x_2)$. Furthermore,
$$\zeta_N=\exp(2\pi i/N)=\exp(2\pi i(u_1/N_2+u_2/N_1))=\zeta_{N_1}^{u_2}\zeta_{N_2}^{u_1}\ ,$$

hence

$$\begin{aligned}\mathfrak{g}(\chi) &= \sum_{x \bmod N} \chi(x)\zeta_N^x \\ &= \sum_{x_1 \bmod N_1,\ x_2 \bmod N_2} \chi_1(x_1)\chi_2(x_2)\zeta_{N_1}^{u_2x_1}\zeta_{N_2}^{u_1x_2} \\ &= \mathfrak{g}(\chi_1; u_2)\mathfrak{g}(\chi_2; u_1) = \chi_1^{-1}(u_2)\chi_2^{-1}(u_1)\mathfrak{g}(\chi_1)\mathfrak{g}(\chi_2)\ ,\end{aligned}$$

so the result follows since $N_2u_2 \equiv 1 \pmod{N_1}$ and $N_1u_1 \equiv 1 \pmod{N_2}$. □

Thanks to the above result, the computation of Gauss sums modulo N can be reduced to the computation of Gauss sums modulo prime powers.

Here a remarkable simplification occurs, due to Odoni: Gauss sums modulo p^a for $a \geq 2$ can be "explicitly computed", in the sense that there is a direct formula not involving a sum over p^a terms for computing them. Although the proof is not difficult, we do not give it, and refer instead to [6] which can be obtained from the author. We use the classical notation $\mathbf{e}(x)$ to mean $e^{2\pi ix}$.

Theorem 6.8 (Odoni et al.)**.** *Let χ be a* primitive *character modulo p^n.*

(1) *Assume that $p \geq 3$ is prime and $n \geq 2$. Write $\chi(1+p) = \mathbf{e}(-b/p^{n-1})$ with $p \nmid b$. Define*

$$A(p) = \frac{p}{\log_p(1+p)} \quad \textit{and} \quad B(p) = A(p)(1 - \log_p(A(p)))\ ,$$

except when $p^n = 3^3$, in which case we define $B(p) = 10$. Then

$$\mathfrak{g}(\chi) = p^{n/2}\mathbf{e}\left(\frac{bB(p)}{p^n}\right)\chi(b) \cdot \begin{cases} 1 & \textit{if } n \geq 2 \textit{ is even,} \\ \left(\dfrac{b}{p}\right) i^{p(p-1)/2} & \textit{if } n \geq 3 \textit{ is odd.}\end{cases}$$

(2) *Let $p = 2$ and assume that $n \geq 4$. Write $\chi(1+p^2) = \mathbf{e}(b/p^{n-2})$ with $p \nmid b$. Define*

$$A(p) = -\frac{p^2}{\log_p(1+p^2)} \quad \textit{and} \quad B(p) = A(p)(1 - \log_p(A(p)))\ ,$$

except when $p^n = 2^4$, in which case we define $B(p) = 13$. Then

$$\mathfrak{g}(\chi) = p^{n/2}\mathbf{e}\left(\frac{bB(p)}{p^n}\right)\chi(b) \cdot \begin{cases} \mathbf{e}\left(\dfrac{b}{8}\right) & \textit{if } n \geq 4 \textit{ is even,} \\ \mathbf{e}\left(\dfrac{(b^2-1)/2+b}{8}\right) & \textit{if } n \geq 5 \textit{ is odd.}\end{cases}$$

(3) *If $p^n = 2^2$, or $p^n = 2^3$ and $\chi(-1) = 1$, we have $\mathfrak{g}(\chi) = p^{n/2}$, and if $p^n = 2^3$ and $\chi(-1) = -1$ we have $\mathfrak{g}(\chi) = p^{n/2}i$.*

Thanks to this theorem, we see that the computation of Gauss sums in the context of Dirichlet characters can be reduced to the computation of Gauss sums modulo p for prime p. This is of course the same as the computation of a Gauss sum for a character of $\mathbb{F}_p^*$.

We recall the available methods for computing a single Gauss sum of this type:

1. The naïve method, time $\widetilde{O}(p)$ (applicable in general, time $\widetilde{O}(N)$).
2. Using the Gross–Koblitz formula, also time $\widetilde{O}(p)$, but the implicit constant is much smaller, and also computations can be done modulo p or p^2 for instance, if desired (applicable only to $N = p$, or in the context of finite fields).
3. Using theta functions, time $\widetilde{O}(p^{1/2})$ (applicable in general, time $\widetilde{O}(N^{1/2})$).

7. Dirichlet L-Series

7.1. **Definition and Main Properties.**

Let χ be a Dirichlet character modulo N. We define the L-function attached to χ as the complex function

$$L(\chi, s) = \sum_{n\ge1} \frac{\chi(n)}{n^s}\ .$$

Since $|\chi(n)| \le 1$, it is clear that $L(\chi, s)$ converges absolutely for $\Re(s) > 1$. Furthermore, since χ is multiplicative, as for the Riemann zeta function we have an *Euler product*

$$L(\chi, s) = \prod_p \frac{1}{1 - \chi(p)/p^s}\ .$$

The denominator of this product being generically of degree 1, this is also called an L-function of degree 1, and conversely, with a suitable definition of the notion of L-function, one can show that these are the only L-functions of degree 1.

If f is the conductor of χ and χ_f is the character modulo f equivalent to χ, it is clear that

$$L(\chi, s) = \prod_{p\mid N,\, p\nmid f} (1 - \chi_f(p)p^{-s})L(\chi_f, s)\ ,$$

so if desired we can always reduce to primitive characters, and this is what we will do in general.

Dirichlet L-series have important analytic and arithmetic properties, some of them conjectural (such as the Riemann Hypothesis), which should (again conjecturally) be shared by all global L-functions, see the discussion in the introduction. We first give the following:

Theorem 7.1. *Let χ be a* primitive *character modulo N, and let $e = 0$ or 1 be such that $\chi(-1) = (-1)^e$.*

(1) *(Analytic continuation.) The function $L(\chi, s)$ can be analytically continued to the whole complex plane into a meromorphic function, which is in fact holomorphic except in the special case $N = 1$, $L(\chi, s) = \zeta(s)$, where it has a unique pole, at $s = 1$, which is simple with residue* 1.

(2) *(Functional equation.) There exists a* functional equation *of the following form: letting $\gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$, we set*

$$\Lambda(\chi, s) = N^{(s+e)/2}\gamma_{\mathbb{R}}(s + e)L(\chi, s) ,$$

where e is as above. Then

$$\Lambda(\chi, 1 - s) = W(\chi)\Lambda(\overline{\chi}, s) ,$$

where $W(\chi)$, the so-called root number*, is a complex number of modulus* 1 *given by the formula $W(\chi) = \mathfrak{g}(\chi)/(i^e N^{1/2})$.*

(3) *(Special values.) For each integer $k \geq 1$ we have the* special values

$$L(\chi, 1 - k) = -\frac{B_k(\chi)}{k} - \delta_{N,1}\delta_{k,1} ,$$

where δ is the Kronecker symbol, and $B_k(\chi)$ are easily computable algebraic numbers. In particular, when $k \not\equiv e \pmod 2$ we have $L(\chi, 1 - k) = 0$ (except when $k = N = 1$).

By the functional equation this is equivalent to the formula for $k \equiv e \pmod 2$, $k \geq 1$:

$$L(\chi, k) = (-1)^{k-1+(k+e)/2}W(\chi)\frac{2^{k-1}\pi^k\overline{B_k(\chi)}}{m^{k-1/2}k!} .$$

7.2. **Computational Issues.** There are several problems that we want to solve, which are best understood in the context of more general L-functions. The first, but not necessarily the most important, is the numerical computation of $L(\chi, s)$ for given χ and s. This problem is of very varying difficulty depending on the size of N, the conductor of χ, and the imaginary part of s (note that if the *real part* of s is quite large, the defining series for $L(\chi, s)$ converges quite well, if not exponentially fast, so there is no problem in that range, and by the functional equation the same is true if the real part of $1 - s$ is quite large).

The problems for $\Im(s)$ large are quite specific, and are already crucial in the case of the Riemann zeta function $\zeta(s)$. It is by an efficient management of this problem (for instance by using the so-called *Riemann–Siegel formula*) that one is able to compute billions of non-trivial zeros of $\zeta(s)$. We will not consider them here, but concentrate on reasonable ranges of s.

The second problem is more specific to general L-functions: in the general situation, we are given an L-function by an Euler product known outside of a finite and small number of "bad primes". Using recipes dating to the late 1960's and well explained in a beautiful paper of Serre [11], one can give the "gamma factor" $\gamma(s)$, and some (but not all) the information about the "conductor", which is the exponential factor.

Let us ignore these problems and assume that we know all the bad primes, gamma factor, conductor, and root number. Note that if we know the gamma factor and the bad primes, using the formulas that we will give below for different values of the argument it is easy to recover the conductor and the root number. What is most difficult to obtain are the Euler factors at the bad primes, and this is the object of current work.

To state the next theorem, which for the moment we state for Dirichlet L-functions, we need still another important special function:

Definition 7.2. *For $x > 0$ we define the* incomplete gamma function $\Gamma(s, x)$ *by*

$$\Gamma(s,x) = \int_x^\infty t^s e^{-t}\,\frac{dt}{t}\ .$$

Note that this integral converges for *all* $s \in \mathbb{C}$, and that it tends to 0 exponentially fast when $x \to \infty$, more precisely $\Gamma(s,x) \sim x^{s-1}e^{-x}$. In addition (but this would carry us too far here) there are many efficient methods to compute it; see however the section on inverse Mellin transforms below.

Theorem 7.3. *Let χ be a* primitive *character modulo N. For all $A > 0$ we have:*

$$\begin{aligned}\Gamma\left(\frac{s+e}{2}\right)L(\chi,s) = \delta_{N,1}\pi^{s/2}&\left(\frac{A^{(s-1)/2}}{s-1}-\frac{A^{s/2}}{s}\right)\\ &+\sum_{n\ge1}\frac{\chi(n)}{n^s}\Gamma\left(\frac{s+e}{2},\frac{\pi n^2A}{N}\right)\\ &+W(\chi)\left(\frac{\pi}{N}\right)^{s-1/2}\sum_{n\ge1}\frac{\overline{\chi}(n)}{n^{1-s}}\Gamma\left(\frac{1-s+e}{2},\frac{\pi n^2}{AN}\right)\ .\end{aligned}$$

Remarks.

(1) Thanks to this theorem, we can compute numerical values of $L(\chi,s)$ (for s in a reasonable range) in time $\widetilde{O}(N^{1/2})$.

(2) The optimal value of A is $A = 1$, but the theorem is stated in this form for several reasons, one of them being that by varying A (for instance taking $A = 1.1$ and $A = 0.9$) one can check the correctness of the implementation, or even compute the root number $W(\chi)$ if it is not known.

(3) To compute values of $L(\chi,s)$ when $\Im(s)$ is large, one does not use the theorem as stated, but variants, see [10].

(4) The above theorem, called the *approximate functional equation*, evidently implies the functional equation itself, so it seems to be more precise; however this is an illusion since one can show that under very mild assumptions functional equations in a large class imply corresponding approximate functional equations.

In fact, let us make this last statement completely precise. For the sake of simplicity we will assume that the L-functions have no poles (this corresponds for Dirichlet L-functions to the requirement that χ not be the trivial character). We begin by the following (where we restrict to certain kinds of gamma products, but it is easy to generalize; incidentally recall the *duplication formula* for the gamma function $\Gamma(s/2)\Gamma((s+1)/2) = 2^{1-s}\pi^{1/2}\Gamma(s)$, which allows the reduction of factors of the type $\Gamma(s+a)$ to several of the type $\Gamma(s/2+a')$ and conversely).

Definition 7.4. *Recall that we have defined $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$, which is the gamma factor attached to L-functions of even characters, for instance to $\zeta(s)$. A* gamma product *is a function of the type*

$$\gamma(s) = f^{s/2}\prod_{1\le i\le d}\Gamma_{\mathbb{R}}(s+b_i)\ ,$$

where $f>0$ is a real number. The number d of gamma factors is called the degree *of $\gamma(s)$.*

Note that the b_i may not be real numbers, but in the case of L-functions attached to motives, they will always be, and in fact be integers.

Proposition 7.5. *Let γ be a gamma product.*

(1) *There exists a function $W(t)$ called the* inverse Mellin transform *of γ such that*

$$\gamma(s) = \int_0^\infty t^s W(t)\,dt/t$$

for $\Re(s)$ sufficiently large (greater than the real part of the rightmost pole of $\gamma(s)$ suffices).

(2) *$W(t)$ is given by the following* Mellin inversion formula *for $t>0$:*

$$W(t) = \mathcal{M}^{-1}(\gamma)(t) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} t^{-s}\gamma(s)\,ds\ ,$$

for any σ larger than the real part of the poles of $\gamma(s)$.

(3) *$W(t)$ tends to 0 exponentially fast when $t\to+\infty$. More precisely, there exist constants A and B (which can easily be made explicit) such that*

$$W(t)\sim At^B\exp(-\pi d(t/f^{1/2})^{2/d})$$

as $t\to\infty$.

Definition 7.6. *Let $\gamma(s)$ be a gamma product and $W(t)$ its inverse Mellin transform. The* incomplete gamma product *$\gamma(s,x)$ is defined for $x>0$ by*

$$\gamma(s,x)=\int_x^\infty t^sW(t)\,\frac{dt}{t}\ .$$

Note that this integral always converges since $W(t)$ tends to 0 exponentially fast when $t\to\infty$. In addition, thanks to the above proposition it is immediate to show that as $x\to\infty$ we have

$$\gamma(s,x)\sim A'x^{B'}\exp(-\pi d(x/f^{1/2})^{2/d})$$

for some other constants A' and B', i.e., with the same exponential decay as $W(t)$.

The main theorem, essentially due to Lavrik, which is an exercise in complex integration is as follows (recall that a function f is of *finite order* $\alpha\ge 0$ if for all $\varepsilon>0$ and sufficiently large $|z|$ we have $|f(z)|\le\exp(|z|^{\alpha+\varepsilon})$):

Theorem 7.7. *For $i=1$ and $i=2$, let $L_i(s)=\sum_{n\ge1}a_i(n)n^{-s}$ be Dirichlet series converging in some right half-plane $\Re(s)\ge\sigma_0$. For $i=1$ and $i=2$, let $\gamma_i(s)$ be gamma products having the same degree d. Assume that the functions $\Lambda_i(s)=\gamma_i(s)L_i(s)$ extend analytically to $\mathbb{C}$ into holomorphic functions of* finite order*, and that we have the functional equation*

$$\Lambda_1(k-s)=w\cdot\Lambda_2(s)$$

for some constant $w\in\mathbb{C}^*$ and some real number k.

Then for all $A>0$, we have

$$\Lambda_1(s)=\sum_{n\ge1}\frac{a_1(n)}{n^s}\gamma_1(s,nA)+w\sum_{n\ge1}\frac{a_2(n)}{n^{k-s}}\gamma_2\Big(k-s,\frac{n}{A}\Big)$$

and symmetrically

$$\Lambda_2(s)=\sum_{n\ge1}\frac{a_2(n)}{n^s}\gamma_2\Big(s,\frac{n}{A}\Big)+w^{-1}\sum_{n\ge1}\frac{a_1(n)}{n^{k-s}}\gamma_1(k-s,nA)\ ,$$

where $\gamma_i(s,x)$ are the corresponding incomplete gamma products.

Note that, as already mentioned, it is immediate to modify this theorem to take into account possible poles of $L_i(s)$.

Since the incomplete gamma products $\gamma_i(s,x)$ tend to 0 exponentially fast when $x\to\infty$, the above formulas are rapidly convergent series. We can make this more precise: if we write as above $\gamma_i(s,x)\sim A_i'x^{B_i'}\exp(-\pi d(x/f_i^{1/2})^{2/d})$, since the convergence of the series is dominated by the exponential term, choosing $A=1$, to have the nth term of the series less than e^{-D}, say, we need (approximately) $\pi d(n/f^{1/2})^{2/d}>D$, in other words $n>(D/(\pi d))^{d/2}f^{1/2}$, with $f=\max(f_1,f_2)$. Thus, if the "conductor" f is large, we may have some trouble. But this stays reasonable for $f<10^8$, say.

The above argument leads to the belief that, apart from special values which can be computed by other methods, the computation of values of L-functions of conductor f requires at least $C \cdot f^{1/2}$ operations. It has however been shown by Hiary (see [8]), that if the "rational part" of f is far from squarefree (for instance if $f = \pi m^3$ for Dirichlet L-functions), the computation can be done faster (in $\widetilde{O}(m)$ in the case $f = \pi m^3$), at least in the case of Dirichlet L-functions. This is perhaps related to Odoni's Theorem 6.8 on the computation of Gauss sums modulo prime powers.

7.3. **Inverse Mellin Transforms.** We thus see that it is necessary to compute inverse Mellin transforms of some common gamma factors. Note that the exponential factors (either involving the conductor and/or π) are easily taken into account: if $\gamma(s) = \mathcal{M}(W)(s) = \int_0^\infty W(t)t^s\,dt/t$ is the Mellin transform of $W(t)$, we have for $a > 0$, setting $u = at$:

$$\int_0^\infty W(at)t^s\,dt/t = \int_0^\infty W(u)u^s a^{-s}\,du/u = a^{-s}\gamma(s)\ ,$$

so the inverse Mellin transform of $a^{-s}\gamma(s)$ is simply $W(at)$.

As we have seen, there exists an explicit formula for the inverse Mellin transform, which is immediate from the Fourier inversion formula. We will see that although this looks quite technical, it is in practice very useful for computing inverse Mellin transforms.

Let us look at the simplest examples (omitting the exponential factors thanks to the above remark):

(1) $\gamma(s) = \Gamma(s/2)$ (occurs for L-functions of even characters, and in particular for $\zeta(s)$). We have $\int_0^\infty e^{-t}t^{s/2}\,dt/t = \Gamma(s/2)$, so setting $t = u^2$ we obtain $\int_0^\infty e^{-u^2}u^s\,2du/u = \Gamma(s/2)$, hence

$$\mathcal{M}^{-1}(\Gamma(s/2)) = 2e^{-x^2}\ .$$

(2) $\gamma(s) = \Gamma((s+1)/2)$ (occurs for L-functions of odd characters). The above formula gives $\int_0^\infty e^{-u^2}u^{s+1}\,2du/u = \Gamma((s+1)/2)$, hence

$$\mathcal{M}^{-1}(\Gamma((s+1)/2)) = 2xe^{-x^2}\ .$$

(3) $\gamma(s) = \Gamma(s)$ (occurs for L-functions attached to modular forms and to elliptic curves). Here we know directly by definition of the gamma function that

$$\mathcal{M}^{-1}(\Gamma(s)) = e^{-x}\ .$$

(4) When there are more Γ-factors, we must not forget to use the duplication formula for the gamma function $\Gamma(s/2)\Gamma((s+1)/2) = 2^{1-s}\Gamma(1/2)\Gamma(s)$. For instance, it is clear that

$$\mathcal{M}^{-1}(\Gamma(s/2)\Gamma((s+1)/2)) = 2\pi^{1/2}e^{-2x}\ .$$

(5) $\gamma(s) = \Gamma(s/2)^2$. Here we introduce a well-known special function, the K-Bessel function. Of course it is just a name, but it can be computed quite efficiently and can be found in many computer algebra packages. We have

$$\mathcal{M}^{-1}(\Gamma(s/2))^2 = 4K_0(2x)$$

(this can be taken as a definition of the function K_0).

(6) $\gamma(s) = \Gamma(s)^2$. From the preceding formula we deduce that

$$\mathcal{M}^{-1}(\Gamma(s)^2) = 2K_0(2x^{1/2}) \ .$$

(7) $\gamma(s) = \Gamma(s)\Gamma(s-1)$. Defining $K_1(x) = -K_0'(x)$, from the preceding formula we deduce that

$$M^{-1}(\Gamma(s)\Gamma(s-1)) = 2K_1(2x^{1/2})/x^{1/2} \ .$$

Exercise: Prove this last formula.

It is clear however that when the gamma factor is more complicated, we cannot write such "explicit" formulas, for instance what must be done for $\gamma(s) = \Gamma(s)\Gamma(s/2)$ or $\gamma(s) = \Gamma(s/2)^3$?. In fact all the above formulas involving K-Bessel functions are "cheats" in the sense that we have simply given a *name* to these inverse Mellin transform, without explaining how to compute them.

However the Mellin inversion formula does provide such a method. The main point to remember (apart of course from the crucial use of the Cauchy residue formula and contour integration), is that the gamma function *tends to zero exponentially fast* on vertical lines, uniformly in the real part (this may seem surprising if you have never seen it since the gamma function grows so fast on the real axis). More precisely, if $\sigma \in \mathbb{R}$ is fixed, then as $|t| \to \infty$ we have precisely

$$|\Gamma(\sigma + it)| \sim |t|^{\sigma-1/2} e^{-\pi|t|/2} (2\pi)^{1/2} \ .$$

This exponential decrease implies that in the Mellin inversion formula we can *shift* the line of integration without changing the value of the integral, as long as we take into account the residues of the poles which are encountered along the way.

The line $\Re(s) = \sigma$ has been chosen so that σ is larger than the real part of any pole of $\gamma(s)$, so shifting to the right does not bring anything. On the other hand, shifting towards the left shows that for any $r < 0$ not a pole of $\gamma(s)$ we have

$$W(t) = \sum_{\substack{s_0 \text{ pole of } \gamma(s) \\ \Re(s_0) > r}} \operatorname{Res}_{s=s_0}(t^{-s}\gamma(s)) + \frac{1}{2\pi i} \int_{r-i\infty}^{r+i\infty} t^{-s}\gamma(s)\, ds \ .$$

Using the reflection formula for the gamma function $\Gamma(s)\Gamma(1-s) = \pi/\sin(s\pi)$, it is easy to show that if r stays say half-way between the real part of two consecutive poles of $\gamma(s)$ then $\gamma(s)$ will tend to 0 exponentially fast on $\Re(s) = r$ as $r \to -\infty$, in other words that the integral tends to 0 (exponentially fast). We thus have the *exact formula*

$$W(t) = \sum_{s_0 \text{ pole of } \gamma(s)} \mathrm{Res}_{s=s_0}(t^{-s}\gamma(s)) \ .$$

Let us see the simples examples of this, taken from those given above.

(1) For $\gamma(s) = \Gamma(s)$ the poles of $\Gamma(s)$ are for $s_0 = -n$, n a positive or zero integer, and since $\Gamma(s) = \Gamma(s+n+1)/((s+n)(s+n-1)\cdots s)$, the residue at $s_0 = -n$ is equal to

$$t^n\Gamma(1)/((-1)(-2)\cdots(-n)) = (-1)^n t^n/n! \ ,$$

so we obtain $W(t) = \sum_{n\ge0}(-1)^n t^n/n! = e^{-t}$. Of course we knew that!

(2) For $\gamma(s) = \Gamma(s)^2$, the inverse Mellin transform is $2K_0(2x^{1/2})$ whose expansion we do *not* yet know. The poles of $\gamma(s)$ are again for $s_0 = -n$, but here all the poles are double poles, so the computation is more complicated. More precisely we have $\Gamma(s)^2 = \Gamma(s+n+1)^2/((s+n)^2(s+n-1)^2\cdots s^2)$, so setting $s = -n+\varepsilon$ with ε small this gives

$$\begin{aligned}\Gamma(-n+\varepsilon)^2 &= \frac{\Gamma(1+\varepsilon)^2}{\varepsilon^2}\frac{1}{(1-\varepsilon)^2\cdots(n-\varepsilon)^2}\\ &= \frac{1+2\Gamma'(1)\varepsilon+O(\varepsilon^2)}{n!^2\varepsilon^2}(1+2\varepsilon/1)(1+2\varepsilon/2)\cdots(1+2\varepsilon/n)\\ &= \frac{1+2\Gamma'(1)\varepsilon+O(\varepsilon^2)}{n!^2\varepsilon^2}(1+2H_n\varepsilon) \ ,\end{aligned}$$

where we recall that $H_n = \sum_{1\le j\le n} 1/j$ is the harmonic sum. Since $t^{-(-n+\varepsilon)} = t^{n-\varepsilon} = t^n(1-\varepsilon\log(t)+O(\varepsilon^2))$, it follows that

$$t^{-(-n+\varepsilon)}\Gamma(-n+\varepsilon)^2 = \frac{t^n}{n!^2\varepsilon^2}(1+\varepsilon(2H_n+2\Gamma'(1)-\log(t))) \ ,$$

so that the residue at $-n$ is equal to $(t^n/n!^2)(2H_n+2\Gamma'(1)-\log(t))$. We thus have $2K_0(2t^{1/2}) = \sum_{n\ge0}(t^n/n!^2)(2H_n+2\Gamma'(1)-\log(t))$, hence using the easily proven fact that $\Gamma'(1) = -\gamma$, where

$$\gamma = \lim_{n\to\infty}(H_n - \log(n)) = 0.57721566490\ldots$$

is Euler's constant, this gives finally the expansion

$$K_0(t) = \sum_{n\ge0}\frac{(t/2)^{2n}}{n!^2}(H_n - \gamma - \log(t/2)) \ .$$

Exercise: In a similar manner, or directly from this formula, find the expansion of $K_1(t)$.

Exercise: Like all inverse Mellin transforms of gamma factors, the function $K_0(x)$ tends to 0 exponentially fast as $x \to \infty$ (more precisely

$K_0(x)\sum(2x/\pi)^{-1/2}e^{-x})$. Note that this is absolutely not "visible" on the expansion given above. Use this remark and the above expansion to write an algorithm which computes Euler's constant γ *very efficiently* to a given accuracy.

It must be remarked that even though the series defining the inverse Mellin transform converge for *all* $x > 0$, one need a large number of terms before the terms become very small when x is large. For instance, we have seen that for $\gamma(s) = \Gamma(s)$ we have $W(t) = \mathcal{M}^{-1}(\gamma)(t) = \sum_{n\geq 0}(-1)^n t^n/n! = e^{-t}$, but this series is not very good for computing e^{-t}.

Exercise: Show that to compute e^{-t} to any reasonable accuracy (even to 1 decimal) we must take at least $n > 3.6 \cdot t$ ($e = 2.718...$), and work to accuracy at most e^{-2t} in an evident sense.

The reason that this is not a good way is that there is catastrophic cancellation in the series. One way to circumvent this problem is to compute e^{-t} as

$$e^{-t} = 1/e^t = 1/\sum_{n\geq 0} t^n/n!\ ,$$

and the cancellation problem disappears. However this is very special to the exponential function, and is not applicable for instance to the K-Bessel function.

Nonetheless, an important result is that for any inverse Mellin transform as above, or more importantly for the corresponding incomplete gamma product, there exist *asymptotic expansions* as $x \to \infty$, in other words nonconvergent series which however give a good approximation if limited to a few terms.

Let us take the simplest example of the incomplete gamma function $\Gamma(s,x) = \int_x^\infty t^s e^{-t}\,dt/t$. The *power series* expansion is easily seen to be (at least for s not a negative or zero integer, otherwise the formula must be slightly modified):

$$\Gamma(s,x) = \Gamma(s) - \sum_{n\geq 0}(-1)^n \frac{x^{n+s}}{n!(s+n)}\ ,$$

which has the same type of (bad when x is large) convergence behavior as e^{-x}. On the other hand, it is immediate to prove by integration by parts that

$$\Gamma(s,x) = e^{-x}x^{s-1}\biggl(1 + \frac{s-1}{x} + \frac{(s-1)(s-2)}{x^2} + \cdots$$
$$+\frac{(s-1)(s-2)\cdots(s-n)}{x^n} + R_n(s,x)\biggr)\ ,$$

and one can show that in reasonable ranges of s and x the modulus of $R_n(s,x)$ is smaller than the first "neglected term" in an evident sense.

This is therefore quite a practical method for computing these functions when x is rather large.

Exercise: Explain why the asymptotic series above terminates when s is a strictly positive integer.

7.4. **Hadamard Products and Explicit Formulas.** This could be the subject of a course in itself, so we will be quite brief. I refer to Mestre's paper [9] for a precise and general statement (note that there are quite a number of evident misprints in the paper).

In Theorem 7.7 we assume that the L-series that we consider satisfy a functional equation, together with some mild growth conditions, in particular that they are of finite order. According to a well-known theorem of complex analysis, this implies that they have a so-called *Hadamard product.* For instance, in the case of the Riemann zeta function, which is of order 1, we have

$$\zeta(s) = \frac{e^{bs}}{s(s-1)\Gamma(s/2)} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{s/\rho} ,$$

where the product is over all nontrivial zeros of $\zeta(s)$ (i.e., such that $0 \le \Re(\rho) \le 1$), and $b = \log(2\pi) - 1 - \gamma$. In fact, this can be written in a much nicer way as follows: recall that $\Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ satisfies $\Lambda(1-s) = \Lambda(s)$. Then

$$s(s-1)\Lambda(s) = \prod_{\rho} \left(1 - \frac{s}{\rho}\right) ,$$

where it is now understood that the product is taken as the limit as $T \to \infty$ of $\prod_{|\Im(\rho)| \le T}(1 - s/\rho)$.

However, almost all L-functions that are used in number theory not only have the above properties, but have also *Euler products.* Taking again the example of $\zeta(s)$, we have for $\Re(s) > 1$ the Euler product $\zeta(s) = \prod_p (1 - 1/p^s)^{-1}$. It follows that (in a suitable range of s) we have equality between two products, hence taking logarithms, equality between two *sums.* In our case the Hadamard product gives

$$\log(\Lambda(s)) = -\log(s(s-1)) + \sum_{\rho} \log(1 - s/\rho) ,$$

while the Euler product gives

$$\begin{aligned}\log(\Lambda(s)) &= -(s/2)\log(\pi) + \log(\Gamma(s/2)) - \sum_{p} \log(1 - 1/p^s) \\ &= -(s/2)\log(\pi) + \log(\Gamma(s/2)) + \sum_{p,k\ge1} 1/(kp^{ks}) ,\end{aligned}$$

Equating the two sides gives a relation between on the one hand a sum over the nontrivial zeros of $\zeta(s)$, and on the other hand a sum over prime powers.

In itself, this is not very useful. The crucial idea is to introduce a test function F which we will choose to the best of our interests, and obtain a formula depending on F and some transforms of it.

This is in fact quite easy to do, and even though not very useful in this case, let us perform the computation for Dirichlet L-function of even primitive characters.

Theorem 7.8. *Let χ be an even primitive Dirichlet character of conductor N, and let F be a real function satisfying a number of easy technical conditions. We have the* explicit formula*:*

$$\begin{aligned}\sum_\rho \Phi(\rho) - 2\delta_{N,1}\int_{-\infty}^{\infty} F(x)\cosh(x/2)\,dx \\ = -\sum_{p,k\geq 1}\frac{\log(p)}{p^{k/2}}(\chi^k(p)F(k\log(p)) + \overline{\chi^k(p)}F(-k\log(p)) \\ + F(0)\log(N/\pi) \\ + \int_0^\infty \left(\frac{e^{-x}}{x}F(0) - \frac{e^{-x/4}}{1-e^{-x}}\frac{F(x/2)+F(-x/2)}{2}\right)dx\ ,\end{aligned}$$

where we set

$$\Phi(s) = \int_{-\infty}^{\infty} F(x)e^{(s-1/2)x}\,dx\ ,$$

and as above the sum on ρ is a sum over all the nontrivial zeros of $L(\chi,s)$ taken symmetrically ($\sum_\rho = \lim_{T\to\infty}\sum_{|\Im(\rho)|\leq T}$).

Remarks.

(1) Write $\rho = 1/2 + i\gamma$ (if the GRH is true all γ are real, but even without GRH we can always write this). Then

$$\Phi(\rho) = \int_{-\infty}^{\infty} F(x)e^{i\gamma x}\,dx = \widehat{F}(\gamma)$$

is simply the value at γ of the *Fourier transform* $\widehat{F}$ of F.

(2) It is immediate to generalize to odd χ or more general L-functions:

Exercise: After studying the proof, generalize to an arbitrary pair of L-functions as in Theorem 7.7.

Proof. The proof is not difficult, but involves a number of integral transform computations. We will omit some detailed justifications which are in fact easy but boring.

As in the theorem, we set

$$\Phi(s) = \int_{-\infty}^{\infty} F(x)e^{(s-1/2)x}\,dx\ ,$$

and we first prove some lemmas.

Lemma 7.9. *We have the inversion formulas valid for any $c > 1$:*

$$F(x) = e^{x/2} \int_{c-i\infty}^{c+i\infty} \Phi(s) e^{-sx}\, ds\ .$$

$$F(-x) = e^{x/2} \int_{c-i\infty}^{c+i\infty} \Phi(1-s) e^{-sx}\, ds\ .$$

Proof. This is in fact a hidden version of the Mellin inversion formula: setting $t = e^x$ in the definition of $\Phi(s)$, we deduce that $\Phi(s) = \int_0^\infty F(\log(t)) t^{s-1/2}\, dt/t$, so that $\Phi(s+1/2)$ is the Mellin transform of $F(\log(t))$. By Mellin inversion we thus have for sufficiently large σ:

$$F(\log(t)) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \Phi(s+1/2) t^{-s}\, ds\ ,$$

so changing s into $s - 1/2$ and t into e^x gives the first formula for $c = \sigma + 1/2$ sufficiently large, and the assumptions on F (which we have not given) imply that we can shift the line of integration to any $c > 1$ without changing the integral.

For the second formula, we simply note that

$$\Phi(1-s) = \int_{-\infty}^{\infty} F(x) e^{-(s-1/2)x}\, dx = \int_{-\infty}^{\infty} F(-x) e^{(s-1/2)x}\, dx\ ,$$

so we simply apply the first formula to $F(-x)$. □

Corollary 7.10. *For any $c > 1$ and any $p \geq 1$ we have*

$$\int_{c-i\infty}^{c+i\infty} \Phi(s) p^{-ks}\, ds = F(k\log(p)) p^{-k/2} \quad \textit{and}$$

$$\int_{c-i\infty}^{c+i\infty} \Phi(1-s) p^{-ks}\, ds = F(-k\log(p)) p^{-k/2}\ .$$

Proof. Simply apply the lemma to $x = k\log(p)$. □

Note that we will also use this corollary for $p = 1$.

Lemma 7.11. *Denote as usual by $\psi(s)$ the logarithmic derivative $\Gamma'(s)/\Gamma(s)$ of the gamma function. We have*

$$\int_{c-i\infty}^{c+i\infty} \Phi(s)\psi(s/2) = \int_0^\infty \left(\frac{e^{-x}}{x} F(0) - \frac{e^{-x/4}}{1-e^{-x}} F(x/2) \right) dx \quad \textit{and}$$

$$\int_{c-i\infty}^{c+i\infty} \Phi(1-s)\psi(s/2) = \int_0^\infty \left(\frac{e^{-x}}{x} F(0) - \frac{e^{-x/4}}{1-e^{-x}} F(-x/2) \right) dx\ .$$

Proof. We use one of the most common integral representations of ψ, see [5]: we have

$$\psi(s) = \int_0^\infty \left(\frac{e^{-x}}{x} - \frac{e^{-sx}}{1-e^{-x}} \right) dx\ .$$

Thus, assuming that we can interchange integrals (which is easy to justify), we have, using the preceding lemma:

$$\begin{aligned}\int_{c-i\infty}^{c+i\infty}\Phi(s)\psi(s/2)\,ds &= \int_0^\infty\left(\frac{e^{-x}}{x}\int_{c-i\infty}^{c+i\infty}\Phi(s)\,ds\right.\\ &\qquad\left.-\frac{1}{1-e^{-x}}\int_{c-i\infty}^{c+i\infty}\Phi(s)e^{-(s/2)x}\,ds\right)\,dx\\ &= \int_0^\infty\left(\frac{e^{-x}}{x}F(0)-\frac{e^{-x/4}}{1-e^{-x}}F(x/2)\right)\,dx\ ,\end{aligned}$$

proving the first formula, and the second follows by changing $F(x)$ into $F(-x)$. □

Proof. (of the theorem). Recall from above that if we set $\Lambda(s) = N^{s/2}\pi^{-s/2}\Gamma(s/2)L(\chi,s)$ we have the functional equation $\Lambda(1-s) = W(\chi)\Lambda(\overline{\chi},s)$ for some $W(\chi)$ of modulus 1.

For $c > 1$, consider the following integral

$$J = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\Phi(s)\frac{\Lambda'(s)}{\Lambda(s)}\,ds\ ,$$

which by our assumptions does not depend on $c > 1$. We shift the line of integration to the left (it is easily seen that this is allowed) to the line $\Re(s) = 1-c$, so by the residue theorem we obtain

$$J = S + \frac{1}{2i\pi}\int_{1-c-i\infty}^{1-c+i\infty}\Phi(s)\frac{\Lambda'(s)}{\Lambda(s)}\,ds\ ,$$

where S is the sum of the residues in the rectangle $[1-c,c]\times\mathbb{R}$. We first have possible poles at $s = 0$ and $s = 1$, which occur only for $N = 1$, and they contribute to S

$$-\delta_{N,1}(\Phi(0)+\Phi(1)) = -2\delta_{N,1}\int_{-\infty}^{\infty}F(x)\cosh(x/2)\,dx\ ,$$

and of course second we have the contributions from the nontrivial zeros ρ, which contribute $\sum_\rho \Phi(\rho)$, where it is understood that zeros are counted with multiplicity, so that

$$S = -2\delta_{N,1}\int_{-\infty}^{\infty}F(x)\cosh(x/2)\,dx + \sum_\rho \Phi(\rho)\ .$$

On the other hand, by the functional equation we have $\Lambda'(1-s)/\Lambda(1-s) = -\overline{\Lambda}'(s)/\overline{\Lambda}(s)$ (note that this does not involve $W(\chi)$), where we write $\overline{\Lambda}(s)$ for $\Lambda(\overline{\chi}, s)$, so that

$$\int_{1-c-i\infty}^{1-c+i\infty} \Phi(s)\frac{\Lambda'(s)}{\Lambda(s)}\,ds = \int_{c-i\infty}^{c+i\infty} \Phi(1-s)\frac{\Lambda'(1-s)}{\Lambda(1-s)}\,ds$$

$$= -\int_{c-i\infty}^{c+i\infty} \Phi(1-s)\frac{\overline{\Lambda}'(s)}{\overline{\Lambda}(s)}\,ds\ .$$

Thus,

$$S = J - \frac{1}{2i\pi}\int_{1-c-i\infty}^{1-c+i\infty} \Phi(s)\frac{\Lambda'(s)}{\Lambda(s)}\,ds$$

$$= \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} \left(\Phi(s)\frac{\Lambda'(s)}{\Lambda(s)} + \Phi(1-s)\frac{\overline{\Lambda}'(s)}{\overline{\Lambda}(s)}\right)\,ds\ .$$

Now by definition we have as above

$$\log(\Lambda(s)) = s/2\log(N/\pi) + \log(\Gamma(s/2)) + \sum_{p,k\geq 1} \chi^k(p)/(kp^{ks})$$

(where the double sum is over primes and integers $k \geq 1$), so

$$\frac{\Lambda'(s)}{\Lambda(s)} = \frac{1}{2}\log(N/\pi) + \frac{1}{2}\psi(s/2) - \sum_{p,k\geq 1} \chi^k(p)\log(p)p^{-ks}\ ,$$

and similarly for $\overline{\Lambda}'(s)/\overline{\Lambda}(s)$. Thus, by the above lemmas and corollaries, we have

$$S = \log(N/\pi)F(0) + J_1 - \sum_{p,k\geq 1}\frac{\log(p)}{p^{k/2}}(\chi^k(p)F(k\log(p)) + \overline{\chi^k(p)}F(-k\log(p)))\ ,$$

where

$$J_1 = \int_0^\infty \left(\frac{e^{-x}}{x}F(0) - \frac{e^{-x/4}}{1-e^{-x}}\frac{F(x/2)+F(-x/2)}{2}\right)\,dx\ ,$$

proving the theorem. □

This theorem can be used in several different directions, and has been an extremely valuable tool in analytic number theory. Just to mention a few:

(1) Since the conductor N occurs, we can obtain *bounds* on N, assuming certain conjectures such as the generalized Riemann hypothesis. For instance, this is how Stark–Odlyzko–Poitou–Serre find *discriminant lower bounds* for number fields. This is also how Mestre finds lower bounds for conductors of abelian varieties, and so on.

(2) When the L-function has a zero at its central point (here of course it usually does not, but for more general L-functions it is important), this can give good upper bounds for the order of the zero.

(3) More generally, suitable choices of the test functions can give information on the nontrivial zeros ρ of small imaginary part.

8. Some Useful Analytic Computational Tools

We finish this course by giving a number of little-known numerical methods which are not always directly related to the computation of L-functions, but which are often very useful.

8.1. **The Euler–MacLaurin Summation Formula.**

This numerical method is *very* well-known (there is in fact even a chapter in Bourbaki devoted to it!), and is as old as Taylor's formula, but deserves to be mentioned since it is very useful. We will be vague on purpose, and refer to [1] or [5] for details. Recall that the *Bernoulli numbers* are defined by the formal power series

$$\frac{T}{e^T-1}=\sum_{n\ge 0}\frac{B_n}{n!}T^n\ .$$

We have $B_0=0$, $B_1=-1/2$, $B_2=1/6$, $B_3=0$, $B_4=-1/30$, and $B_{2k+1}=0$ for $k\ge 1$.

Let f be a C^∞ function defined on $\mathbb{R}>0$. The basic statement of the Euler–MacLaurin formula is that there exists a constant $z=z(f)$ such that

$$\sum_{n=1}^{N}f(n)=\int_1^N f(t)\,dt+z(f)+\frac{f(N)}{2}+\sum_{1\le k\le p}\frac{B_{2k}}{(2k)!}f^{(2k-1)}(N)+R_p(N)\ ,$$

where $R_p(N)$ is "small", in general smaller than the first neglected term, as in most asymptotic series.

The above formula can be slightly modified at will, first by changing the lower bound of summation and/or of integration (which simply changes the constant $z(f)$), and second by writing $\int_1^N f(t)\,dt+z(f)=z'(f)-\int_N^\infty f(t)\,dt$ (when f tends to 0 sufficiently fast for the integral to converge), where $z'(f)=z(f)+\int_1^\infty f(t)\,dt$.

The Euler–MacLaurin summation formula can be used in many contexts, but we mention the two most important ones.

• First, to have some idea of the size of $\sum_{n=1}^N f(n)$. Let us take an example. Consider $S_2(N)=\sum_{n=1}^N n^2\log(n)$. Note incidentally that

$$\exp(S_2(N))=\prod_{n=1}^{N}n^{n^2}=1^{1^2}2^{2^2}\cdots N^{N^2}\ .$$

What is the size of this generalized kind of factorial? Euler–MacLaurin tells us that there exists a constant z such that

$$S_2(N)=\int_1^N t^2\log(t)\,dt+z+\frac{N^2\log(N)}{2}$$
$$+\frac{B_2}{2!}(N^2\log(N))'+\frac{B_4}{4!}(N^2\log(N))'''+\cdots\ .$$

We have $\int_1^N t^2\log(t)\,dt=(N^3/3)\log(N)-(N^3-1)/9$, $(N^2\log(N))'=2N\log(N)+N$, $(N^2\log(N))''=2\log(N)+3$, and $(N^2\log(N))'''=2/N$, so using $B_2=1/6$ we obtain for some other constant z':

$$S_2(N)=\frac{N^3\log(N)}{3}-\frac{N^3}{9}+\frac{N^2\log(N)}{2}+\frac{N\log(N)}{6}+\frac{N}{12}+z'+O\left(\frac{1}{N}\right)\ ,$$

which essentially answers our question, up to the determination of the constant z'. Thus we obtain a generalized Stirling's formula:

$$\exp(S_2(N))=N^{N^3/3+N^2/2+N/6}e^{-(N^3/9-N/12)}C\ ,$$

where $C=\exp(z')$ is an a priori unknown constant. In the case of the usual Stirling's formula we have $C=(2\pi)^{1/2}$, so we can ask for a similar formula here. And indeed, such a formula exists: we have

$$C=\exp(\zeta(3)/(4\pi^2))\ .$$

Exercise: Do a similar (but simpler) computation for $S_1(N)=\sum_{1\le n\le N}n\log(n)$. The corresponding constant is explicit but more difficult (it involves $\zeta'(-1)$; more generally the constant in $S_r(N)$ involves $\zeta'(-r)$).

• The second use of the Euler–MacLaurin formula is to increase considerably the speed of convergence of slowly convergent series. For instance, if you want to compute $\zeta(3)$ directly using the series $\zeta(3)=\sum_{n\ge1}1/n^3$, since the remainder term after N terms is asymptotic to $1/(2N^2)$ you will never get more than 15 or 20 decimals of accuracy. On the other hand, it is immediate to use Euler–MacLaurin:

Exercise: Write a computer program implementing the computation of $\zeta(3)$ (and more generally of $\zeta(s)$ for reasonable s) using Euler–MacLaurin, and compute it to 100 decimals.

A variant of the method is to compute limits: a typical example is the computation of Euler's constant

$$\gamma=\lim_{N\to\infty}\left(\sum_{n=1}^N\frac{1}{n}-\log(N)\right)\ .$$

Using Euler–MacLaurin, it is immediate to find the *asymptotic expansion*

$$\sum_{n=1}^{N} \frac{1}{n} = \log(N) + \gamma + \frac{1}{2N} - \sum_{k\ge 1} \frac{B_{2k}}{2kN^{2k}}$$

(note that this is not a misprint, the last denominator is $2kN^{2k}$, not $(2k)!N^{2k}$).

Exercise: Implement the above, and compute γ to 100 decimal digits.

Note that this is *not* the fastest way to compute Euler's constant, the method using Bessel functions given above is better.

8.2. **Zagier's Extrapolation Method.** The following nice trick is due to D. Zagier. Assume that you have a sequence u_n that you suspect of converging to some limit a_0 when $n \to \infty$ in a regular manner. How do you give a reasonable numerical estimate of a_0 ?

Assume for instance that as $n \to \infty$ we have $u_n = \sum_{0\le i\le p} a_i/n^i + O(n^{-p-1})$ for any p. One idea would be to choosing for n suitable values and solve a linear system. This would in general be quite unstable and inaccurate. Zagier's trick is instead to proceed as follows: choose some reasonable integer k, say $k = 10$, set $v_n = n^k u_n$, and compute the kth *forward difference* $\Delta^k(v_n)$ of this sequence (the forward difference of a sequence w_n is the sequence $\Delta(w)_n = w_{n+1} - w_n$). Note that

$$v_n = a_0 n^k + \sum_{1\le i\le k} a_i n^{k-i} + O(1/n)\ .$$

The two crucial points are the following:

- The kth forward difference of a polynomial of degree less than or equal to $k-1$ vanishes, and that of n^k is equal to $k!$.
- Assuming reasonable regularity conditions, the kth forward difference of an asymptotic expansion beginning at $1/n$ will begin at $1/n^{k+1}$.

Thus, under reasonable assumptions we have

$$a_0 = \Delta^k(v)_n/k! + O(1/n^{k+1})\ ,$$

so choosing n large enough can give a good estimate for a_0.

A number of remarks concerning this basic method:

Remarks

1. It is usually preferable to apply this not to the sequence u_n itself, but for instance to the sequence u_{100n}, if it is not too expensive to compute.
2. It is immediate to modify the method to compute further coefficients a_1, a_2, etc...
3. If the asymptotic expansion of u_n is in powers of $1/n^{1/2}$, say, simply apply the method to the sequence u_{n^2} or u_{100n^2}.

Example. Let us compute numerically the constant occurring in the first example of the use of Euler–MacLaurin that we have given. We set

$$u_N = \sum_{1\le n\le N} n^2\log(n) - (N^3/3+N^2/2+N/6)\log(N) + N^3/9 - N/12\ .$$

We compute for instance that $u_{1000} = 0.0304456\cdots$, which has only 4 correct decimal digits. On the other hand, if we apply the above trick with $k = 12$ and $N = 100$, we find

$$a_0 = \lim_{N\to\infty} u_N = 0.0304484570583932707802515304696767\cdots$$

with 29 correct decimal digits (recall that the exact value is $\zeta(3)/(4\pi^2) = 0.03044845705839327078025153047115477\cdots$).

8.3. **Computation of Euler Sums and Euler Products.** Assume that we want to compute numerically

$$S_1 = \prod_p \left(1+\frac{1}{p^2}\right)\ ,$$

where here and elsewhere, the expression $\prod_p$ always means the product over all prime numbers. Trying to compute it using a large table of prime numbers will not give much accuracy: if we use primes up to X, we will make an error of the order of $1/X$, so it will be next to impossible to have more than 8 or 9 decimal digits.

On the other hand, if we simply notice that $1+1/p^2 = (1-1/p^4)/(1-1/p^2)$, by definition of the Euler product for the Riemann zeta function this implies that

$$S_2 = \frac{\zeta(2)}{\zeta(4)} = \frac{\pi^2/6}{\pi^4/90} = \frac{15}{\pi^2} = 1.519817754635066571658\cdots$$

Unfortunately this is based on a special identity. What if we wanted instead to compute $S_2 = \prod_p(1+2/p^2)$? There is no special identity to help us here.

The way around this problem is to approximate the function of which we want to take the product (here $1+2/p^2$) by *infinite products* of values of the Riemann zeta function. Let us do it step by step before giving the general formula.

When p is large, $1+2/p^2$ is close to $1/(1-1/p^2)^2$, which is the Euler factor for $\zeta(2)^2$. More precisely, $(1+2/p^2)(1-1/p^2)^2 = 1-3/p^4+2/p^6$, so we deduce that

$$S_2 = \zeta(2)^2\prod_p(1-3/p^4+2/p^6) = (\pi^4/36)\prod_p(1-3/p^4+2/p^6)\ .$$

Even though this looks more complicated, what we have gained is that the new Euler product converges *much* faster. Once again, if we compute it for p up to 10^8, say, instead of having 8 decimal digits we now have approximately 24 decimal digits (convergence in $1/X^3$ instead of $1/X$). But there is no reason to stop there: we have $(1-3/p^4+2/p^6)/(1-1/p^4)^3 = 1+O(1/p^6)$ with evident notation and explicit formulas if desired, to we get an even better approximation by writing $S_2 = \zeta(2)^2/\zeta(4)^3\prod_p(1+O(1/p^6))$, with convergence in $1/X^5$. More generally, it is easy to compute by induction exponents $a_n \in \mathbb{Z}$ such that $S_2 = \prod_{2\le n\le N}\zeta(n)^{a_n}\prod_p(1+O(1/p^{N+1}))$ (in our case $a_n = 0$ for n odd but this will not be true in general). It can be shown in essentially all examples that one can pass to the limit, and for instance here write $S_2 = \prod_{n\ge2}\zeta(n)^{a_n}$.

Exercise:

(1) Compute explicitly the recursion for the a_n in the example of S_2.

(2) More generally, if $S = \prod_p f(p)$, where $f(p)$ has a convergent series expansion in $1/p$ starting with $f(p) = 1+1/p^b+o(1/p^b)$ with $b > 1$ (not necessarily integral), express S as a product of zeta values raised to suitable exponents, and find the recursion for these exponents.

An important remark needs to be made here: even though the product $\prod_{n\ge2}\zeta(n)^{a_n}$ may be convergent, it may converge rather slowly: remember that when n is large we have $\zeta(n)-1\sim1/2^n$, so that in fact if the a_n grow like 3^n the product will not even converge. The way around this, which must be used even when the product converges, is as follows: choose a reasonable integer N, for instance $N = 50$, and compute $\prod_{p\le50}f(p)$, which is of course very fast. Then the tail $\prod_{p>50}f(p)$ of the Euler product will be equal to $\prod_{n\ge2}\zeta_{>50}(n)^{a_n}$, where $\zeta_{>N}(n)$ is the zeta function without its Euler factors up to N, in other words $\zeta_{>N}(n) = \zeta(n)\prod_{p\le N}(1-1/p^n)$ (I am assuming here that we have zeta values at integers as in the S_2 example above, but it is immediate to generalize). Since $\zeta_{>N}(n)-1\sim1/(N+1)^n$, the convergence of our zeta product will of course be considerably faster.

Finally, note that by using the power series expansion of the logarithm together with *Möbius inversion*, it is immediate to do the same for Euler *sums*, for instance to compute $\sum_p 1/p^2$ and the like, see [5] for details.

8.4. **Summation of Alternating Series.** This is due to Rodriguez–Villegas, Zagier, and the author.

We have seen above the use of the Euler–MacLaurin summation formula to sum quite general types of series. If the series is *alternating* (the terms alternate in sign), the method cannot be used as is, but it is trivial to modify it: simply write

$$\sum_{n\ge1}(-1)^nf(n) = \sum_{n\ge1}f(2n) - \sum_{n\ge1}f(2n-1)$$

and apply Euler–MacLaurin to each sum. One can even do better and avoid this double computation, but this is not what I want to mention here.

A completely different method which is much simpler since it avoids completely the computation of derivatives and Bernoulli numbers, due to the above authors, is as follows. The idea is to express (if possible) $f(n)$ as a *moment*

$$f(n) = \int_0^1 x^n w(x)\,dx$$

for some *weight function* $w(x)$. Then it is clear that

$$S = \sum_{n\ge 0}(-1)^n f(n) = \int_0^1 \frac{1}{1+x} w(x)\,dx\ .$$

Assume that $P_n(X)$ is a polynomial of degree n such that $P_n(-1) \neq 0$. Evidently

$$\frac{P_n(-1) - P_n(-1)}{X+1} = \sum_{k=0}^{n-1} c_{n,k} X^k$$

is still a polynomial (of degree $n-1$), and we note the trivial fact that

$$\begin{aligned} S &= \frac{1}{P_n(-1)} \int_0^1 \frac{P_n(-1)}{1+x} w(x)\,dx \\ &= \frac{1}{P_n(-1)} \left(\int_0^1 \frac{P_n(-1) - P_n(x)}{1+x} w(x)\,dx + \int_0^1 \frac{P_n(x)}{1+x} w(x)\,dx \right) \\ &= \frac{1}{P_n(-1)} \sum_{k=0}^{n-1} c_{n,k} f(k) + R_n\ , \end{aligned}$$

with

$$|R_n| \le \frac{M_n}{|P_n(-1)|} \int_0^1 \frac{1}{1+x} w(x)\,dx = \frac{M_n}{|P_n(-1)|} S\ ,$$

and where $M_n = \sup_{x\in[0,1]} |P_n(x)|$. Thus if we can manage to have $M_n/|P_n(-1)|$ small, we obtain a good approximation to S.

It is a classical result that the best choice for P_n are the shifted Chebychev polynomials defined by $P_n(\sin^2(t)) = \cos(2nt)$, but in any case we can use these polynomials and ignore that they are the best.

This leads to an incredibly simple algorithm which we write explicitly:

```
$d \leftarrow (3+\sqrt{8})^n$; $d \leftarrow (d+1/d)/2$; $b \leftarrow -1$; $c \leftarrow -d$; $s \leftarrow 0$; For $k = 0, \dots, n-1$ do:
    $c \leftarrow b - c$; $s \leftarrow s + c \cdot f(k)$; $b \leftarrow (k+n)(k-n)b/((k+1/2)(k+1))$;
    The result is $s/d$.
```

The convergence is in 5.83^{-n}.

It is interesting to note that, even though this algorithm is designed to work with functions f of the form $f(n) = \int_0^1 x^n w(x)\,dx$ with w continuous and positive, it is in fact valid in regions where its validity is not only not proved but even false. For example:

Exercise: It is well-known that the Riemann zeta function $\zeta(s)$ can be extended analytically to the whole complex plane, and that we have for instance $\zeta(-1) = -1/12$ and $\zeta(-2) = 0$. Apply the above algorithm to the *alternating* zeta function

$$\beta(s) = \sum_{n\geq 1}(-1)^{n-1}\frac{1}{n^s} = \left(1 - \frac{1}{2^{s-1}}\right)\zeta(s)$$

(incidentally, prove this identity), and by using the above algorithm, show the nonconvergent "identities"

$$1 - 2 + 3 - 4 + \cdots = 1/4 \quad\text{and}\quad 1 - 2^2 + 3^2 - 4^2 + \cdots = 0\ .$$

8.5. **Numerical Differentiation.** The problem is as follows: given a function f, say defined and C^∞ on a real interval, compute $f'(x_0)$ for a given value of x_0. To be able to analyze the problem, we will assume that $f'(x_0)$ is not too close to 0, and that we want to compute it to a given *relative accuracy*, which is what is usually required in numerical analysis.

The naïve, although reasonable, approach, is to choose a small $h > 0$ and compute $(f(x_0+h)-f(x_0))/h$. However, it is clear that (using the same number of function evaluations) the formula $(f(x_0+h)-f(x_0-h))/(2h)$ will be better. Let us analyze this in detail. For simplicity we will assume that all the derivatives of f around x_0 that we consider are neither too small nor too large in absolute value. It is easy to modify the analysis to treat the general case.

Assume f computed to a relative accuracy of ε, in other words that we know values $\tilde{f}(x)$ such that $\tilde{f}(x)(1-\varepsilon) < f(x) < \tilde{f}(x)(1+\varepsilon)$ (the inequalities being reversed if $f(x) < 0$). The absolute error in computing $(f(x_0+h)-f(x_0-h))/(2h)$ is thus essentially equal to $\varepsilon|f(x_0)|/h$. On the other hand, by Taylor's theorem we have $(f(x_0+h)-f(x_0-h))/(2h) = f'(x_0) + (h^2/6)f'''(x)$ for some x close to x_0, so the absolute error made in computing $f'(x_0)$ as $(f(x_0+h)-f(x_0-h))/(2h)$ is close to $\varepsilon|f(x_0)|/h + (h^2/6)|f'''(x_0)|$. For a given value of ε (i.e., the accuracy to which we compute f) the optimal value of h is $(3\varepsilon|f(x_0)/f'''(x_0)|)^{1/3}$ for an absolute error of $(1/2)(3\varepsilon|f(x_0)f'''(x_0)|)^{2/3}$ hence a relative error of $(3\varepsilon|f(x_0)f'''(x_0)|)^{2/3}/(2|f'(x_0)|)$.

Since we have assumed that the derivatives have reasonable size, the relative error is roughly $C\varepsilon^{2/3}$, so if we want this error to be less than η, say, we need ε of the order of $\eta^{3/2}$, and h will be of the order of $\eta^{1/2}$.

Note that this result is not completely intuitive. For instance, assume that we want to compute derivatives to 38 decimal digits. With our assumptions, we choose h around 10^{-19}, and perform the computations with 57 decimals of relative accuracy. If for some reason or other we are limited to 38 decimals in the computation of f, the "intuitive" way would be also to choose $h = 10^{-19}$, and the above analysis shows that we would obtain only approximately 19 decimals. On the other hand, if we chose $h = 10^{-13}$ for instance, close to $10^{-38/3}$, we would obtain 25 decimals.

There are of course many other formulas for computing $f'(x_0)$, or for computing higher derivatives, which can all easily be analyzed as above. For instance (exercise), one can look for approximations to $f'(x_0)$ of the form $S = (\sum_{1\le i\le 3} \lambda_i f(x_0 + h/a_i))/h$, for any nonzero and pairwise distinct a_i, and we find that this is possible as soon as $\sum_{1\le i\le 3} a_i = 0$ (for instance, if $(a_1, a_2, a_3) = (-3, 1, 2)$ we have $(\lambda_1, \lambda_2, \lambda_3) = (-27, -5, 32)/20$), and the absolute error is then of the form $C_1/h + C2h^3$, so the same analysis shows that we should work with accuracy $\varepsilon^{4/3}$ instead of $\varepsilon^{3/2}$. Even though we have $3/2$ times more evaluations of f, we require less accuracy: for instance, if f requires time $O(D^a)$ to be computed to D decimals, as soon as $(3/2) \cdot ((4/3)D)^a < ((3/2)D)^a$, i.e., $3/2 < (9/8)^a$, hence $a \ge 3.45$, this new method will be faster.

Perhaps the best known method with more function evaluations is the approximation

$$f'(x_0) \approx (f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h))/(12h)\ ,$$

which requires accuracy $\varepsilon^{5/4}$, and since this requires 4 evaluations of f, this is faster than the first method as soon as $2 \cdot (5/4)^a < (3/2)^a$, in other words $a > 3.81$, and faster than the second method as soon as $(4/3) \cdot (5/4)^a < (4/3)^a$, in other words $a > 4.46$. To summarize, use the first method if $a < 3.45$, the second method if $3.45 \le a < 4.46$, and the third if $a > 4.46$. Of course this game can be continued at will, but there is not much point in doing so. In practice the first method is sufficient.

8.6. **Double Exponential Numerical Integration.** A remarkable although little-known technique invented around 1970 deals with *numerical integration* (the numerical computation of a definite integral $\int_a^b f(t)\,dt$, where a and b may even be $\pm\infty$). In usual numerical analysis courses one teaches very elementary techniques such as the trapezoidal rule, Simpson's rule, or more sophisticated methods such as Romberg or Gaussian integration. These methods apply to very general classes of functions $f(t)$, but are unable to compute more than a few decimal digits of the result.

However, in most mathematical (as opposed for instance to physical) contexts, the function $f(t)$ is *extremely regular*, typically holomorphic or meromorphic, at least in some domain of the complex plane. It was observed in the late 1960's by two Japanese mathematicians Takahashi and Mori, that this property can be used to obtain a *very simple* and *incredibly accurate* method to compute definite integrals of such functions. It is now instantaneous to compute 100 decimal digits (out of the question for classical methods), and takes only a few seconds to compute 500 decimal digits, say.

In view of its importance it is essential to have some knowledge of this method. It can of course be applied in a wide variety of contexts, but note also that in his thesis, P. Molin has applied it specifically to the *rigorous* and *practical* computation of values of L-functions, which brings us back to our main theme.

There are two basic ideas behind this method. The first is in fact a theorem, which I state in a vague form: If F is a holomorphic function which tends to 0 "sufficiently fast" when $x \to \pm\infty$, x real, then the most efficient method to compute $\int_{\mathbb{R}} F(t)\,dt$ is indeed the trapezoidal rule. Note that this is a *theorem*, not so difficult but a little surprising nonetheless. The definition of "sufficiently fast" can be made precise. In practice, it means at least like e^{-ax^2} ($e^{-a|x|}$ is not fast enough), but it can be shown that the best results are obtained with functions tending to 0 *doubly exponentially fast* such as $\exp(-\exp(a|x|))$. Note that it would be (very slightly) worse to choose functions tending to 0 even faster.

To be more precise, we have an estimate coming for instance from the *Euler–MacLaurin summation formula*:

$$\int_{-\infty}^{\infty} F(t)\,dt = h \sum_{n=-N}^{N} F(nh) + R_N(h) \ ,$$

and under suitable holomorphy conditions on F, if we choose $h = a\log(N)/N$ for some constant a close to 1, the remainder term $R_N(h)$ will satisfy $R_n(h) = O(e^{-bN/\log(N)})$ for some other (reasonable) constant b, showing exponential convergence of the method.

The second and of course crucial idea of the method is as follows: evidently not all functions are doubly-exponentially tending to 0 at $\pm\infty$, and definite integrals are not all from $-\infty$ to $+\infty$. But it is possible to reduce to this case by using clever *changes of variable* (the essential condition of holomorphy must of course be preserved).

Let us consider the simplest example, but others that we give below are variations on the same idea. Assume that we want to compute

$$I = \int_{-1}^{1} f(x)\,dx \ .$$

We make the "magical" change of variable $x = \phi(t) = \tanh(\sinh(t))$, so that if we set $F(t) = f(\phi(t))$ we have

$$I = \int_{-\infty}^{\infty} F(t)\phi'(t)\,dt\ .$$

Because of the elementary properties of the hyperbolic sine and tangent, we have gained two things at once: first the integral from -1 to 1 is now from $-\infty$ to ∞, but most importantly the function $\phi'(t)$ is easily seen to tend to 0 doubly exponentially. We thus obtain an *exponentially good approximation*

$$\int_{-1}^{1} f(x)\,dx = h\sum_{n=-N}^{N} f(\phi(nh))\phi'(nh) + R_N(h)\ .$$

To give an idea of the method, if one takes $h = 1/200$ and $N = 500$, hence only 1000 evaluations of the function f, one can compute I to several hundred decimal places!

Before continuing, I would like to comment that in this theory many results are not completely rigorous: the method works very well, but the proof that it does is sometimes missing. Thus I cannot resist giving a *proven and precise* theorem due to P. Molin (which is of course just an example). We keep the above notation $\phi(t) = \tanh(\sinh(t))$, and note that $\phi'(t) = \cosh(t)/\cosh^2(\sinh(t))$.

Theorem 8.1 (Molin)**.** *Let f be holomorphic on the disc $D = D(0,2)$ centered at the origin and of radius 2. Then for all $N \geq 1$, if we choose $h = \log(5N)/N$ we have*

$$\int_{-1}^{1} f(x)\,dx = h\sum_{n=-N}^{N} f(\phi(nh))\phi'(nh) + R_N\ ,$$

where

$$|R_N| \leq \left(e^4 \sup_D |f|\right)\exp(-5N/\log(5N))\ .$$

Coming back to the general situation, I briefly comment on the computation of general definite integrals $\int_a^b f(t)\,dt$.

(1) If a and b are finite, we can reduce to $[-1,1]$ by affine changes of variable.

(2) If a (or b) is finite and the function has an algebraic singularity at a (or b), we remove the singularity by a polynomial change of variable.

(3) If $a = 0$ (say) and $b = \infty$, then if f does *not* tend to 0 exponentially fast (for instance $f(x) \sim 1/x^k$), we use $x = \phi(t) = \exp(\sinh(t))$.

(4) If $a = 0$ (say) and $b = \infty$ and if f does tend to 0 exponentially fast (for instance $f(x) \sim e^{-ax}$ or $f(x) \sim e^{-ax^2}$), we use $x = \phi(t) = \exp(t - \exp(-t))$.

(5) If $a = -\infty$ and $b = \infty$, use $x = \phi(t) = \sinh(\sinh(t))$ if f does not tend to 0 exponentially fast, and $x = \phi(t) = \sinh(t)$ otherwise.

The problem of *oscillating* integrals such as $\int_0^\infty f(x)\sin(x)\,dx$ is more subtle, but there does exist similar methods when, as here, the oscillations are completely under control.

Remark. The theorems are valid when the function is holomorphic in a sufficiently large region compared to the path of integration. If the function is only *meromorphic*, with known poles, the direct application of the formulas may give totally wrong answers. However, if we take into account the poles, we can recover perfect agreement. Example of bad behavior: $f(t) = 1/(1+t^2)$ (poles $\pm i$). Integrating on the intervals $[0, \infty]$, $[0, 1000]$, or even $[-\infty, \infty]$, which involve different changes of variables, give perfect results (the latter being somewhat surprising). On the other hand, integrating on $[-1000, 1000]$ gives a totally wrong answer because the poles are "too close", but it is easy to take them into account if desired.

Apart from the above pathological behavior, let us give a couple of examples where we must slightly modify the direct use of doubly-exponential integration techniques.

- Assume for instance that we want to compute

$$J = \int_1^\infty \left(\frac{1+e^{-x}}{x}\right)^2\,dx\ ,$$

and that we use the built-in function `intnum` of `Pari/GP` for doing so. The function tends to 0 slowly at infinity, so we should compute it using the `GP` syntax `[1]` to represent ∞, so we write `f(x)=((1+exp(-x))/x)^2;`, then `intnum(x=1,[1],f(x))`. This will give some sort of error, because the software will try to evaluate $\exp(-x)$ for large values of x, which it cannot do since there is exponent underflow. To compute the result, we need to split it into its slow part and fast part: when a function tends exponentially fast to 0, ∞ is represented as `[[1],1]`, so we write $J = J_1 + J_2$, with J_1 and J_2 computed by:
`J1=intnum(x=1,[[1],1],(exp(-2*x)+2*exp(-x))/x^2);` and
`J2=intnum(x=1,[1],1/x^2);` (which of course is equal to 1), giving

$$J = 1.3345252753723345485962398139190637\cdots\ .$$

Note that we could have tried to "cheat" and written directly
`intnum(x=1,[[1],1],f(x))`, but the answer would be wrong, because the software would have assumed that $f(x)$ tends to 0 exponentially fast, which is not the case.

• A second situation where we must be careful is when we have "apparent singularities" which are not real singularities. Consider the function $f(x) = (\exp(x) - 1 - x)/x^2$. It has an apparent singularity at $x = 0$ but in fact it is completely regular. If you ask `J=intnum(x=0,1,f(x))`, you will get a result which is reasonably correct, but never more than 19 decimals, say. The reason is *not* due to a defect in the numerical integration routine, but more in the computation of $f(x)$: if you simply write `f(x)=(exp(x)-1-x)/x^2;`, the results will be bad for x close to 0.

Assuming that you want 38 decimals, say, the solution is to write `f(x)=if(x<10^(-10),1/2+x/6+x^2/24+x^3/120,(exp(x)-1-x)/x^2);` and now we obtain the value of our integral as

$$J = 0.59962032299535865949972137289656934022\cdots$$

8.7. **The Use of Abel–Plana for Definite Summation.** We finish this course by describing an identity, which is perhaps not so computationally useful, but which is quite amusing. Consider for instance the following theorem:

Theorem 8.2. *Define by convention* $\sin(n/10)/n$ *as equal to its limit* $1/10$ *when* $n = 0$*, and define* $\sum'_{n\ge0} f(n)$ *as* $f(0)/2 + \sum_{n\ge1} f(n)$*. We have*

$$\sum_{n\ge0}{}' \left(\frac{\sin(n/10)}{n}\right)^k = \int_0^\infty \left(\frac{\sin(x/10)}{x}\right)^k$$

for $1 \le k \le 62$*, but not for* $k \ge 63$*.*

If you do not like all these conventions, replace the left-hand side by

$$\frac{1}{2\cdot 10^k} + \sum_{n\ge1} \left(\frac{\sin(n/10)}{n}\right)^k .$$

It is clear that something is going on: it is the Abel–Plana formula. There are several forms of this formula, here is one of them:

Theorem 8.3 (Abel–Plana)**.** *Assume that* f *is an entire function and that* $f(z) = o(\exp(2\pi|\Im(z)|))$ *as* $|\Im(z)| \to \infty$ *uniformly in vertical strips of bounded width, and a number of less important additional conditions which we omit. Then*

$$\begin{aligned}\sum_{m\ge1} f(m) &= \int_0^\infty f(t)\,dt - \frac{f(0)}{2} + i\int_0^\infty \frac{f(it)-f(-it)}{e^{2\pi t}-1}\,dt\\ &= \int_{1/2}^\infty f(t)\,dt - i\int_0^\infty \frac{f(1/2+it)-f(1/2-it)}{e^{2\pi t}+1}\,dt\ .\end{aligned}$$

In particular, if the function f *is* even, *we have*

$$\frac{f(0)}{2} + \sum_{m\ge1} f(m) = \int_0^\infty f(t)\,dt\ .$$

Since we have seen above that using doubly-exponential techniques it is easy to compute numerically a definite *integral*, the Abel–Plana formula can be used to compute numerically a *sum*. Note that in the first version of the formula there is an apparent singularity (but which is not a singularity) at $t = 0$, and the second version avoids this problem.

References

[1] N. Bourbaki, *Dvéloppement tayloriens généralisés. Formule sommatoire d'Euler–MacLaurin*, Fonctions d'une variable réelle, Chap. 6.

[2] H. Cohen, *A Course in Computational Algebraic Number Theory (fourth corrected printing)*, Graduate Texts in Math. **138**, Springer-Verlag, 2000.

[3] H. Cohen, *Advanced Topics in Computational Number Theory*, Graduate Texts in Math. **193**, Springer-Verlag, 2000.

[4] H. Cohen, *Number Theory I, Tools and Diophantine Equations*, Graduate Texts in Math. **239**, Springer-Verlag, 2007.

[5] H. Cohen, *Number Theory II, Analytic and Modern Tools*, Graduate Texts in Math. **240**, Springer-Verlag, 2007.

[6] H. Cohen, *A p-adic stationary phase theorem and applications*, preprint.

[7] H. Cohen and D. Zagier, *Vanishing and nonvanishing theta values*, Ann. Sci. Math. Quebec, to appear.

[8] G. Hiary, *Computing Dirichlet character sums to a power-full modulus*, ArXiv preprint 1205.4687v2.

[9] J.-F. Mestre, *Formules explicites et minorations de conducteurs de variétés algébriques*, Compositio Math. **58** (1986). pp. 209–232.

[10] M. Rubinstein, *Computational methods and experiments in analytic number theory*, In: Recent Perspectives in Random Matrix Theory and Number Theory, F. Mezzadri and N. Snaith, eds (2005), pp. 407–483.

[11] J.-P. Serre, *Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures)*, Séminaire Delange–Pisot–Poitou **11** (1969–1970), exp. 19, pp. 1–15.