JUIN / JUNE 2002

IN THIS ISSUE / DANS CE NUMÉRO

24
No 2

Published by / Publié par
The Academy of Science of the Royal Society of Canada
L'Académie des sciences de la Société royale du Canada
c/o Department of Mathematics and Statistics
Carleton University

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 33–54

RECENT DEVELOPMENTS IN THE LANGLANDS PROGRAM

M. RAM MURTY, FRSC

RÉSUMÉ. Cette article est une exposition des résultats récents dans le programme de Langlands. Notamment, c'est un exposé du travail de Shahidi et Kim sur la théorie des fonctions L attachées aux puissances symétriques tensorielles d'une représentation automorphe de GL_n.

This article is based on a a series of lectures delivered at the Centre de Recherches Mathématiques in Montréal in April 2002. These lectures focussed solely on recent progress in the Langlands program over number fields. Spectacular progress has been made recently in the function field case but we do not report on that here.

We will begin with an informal motivation discussing four outstanding conjectures (still unresolved) that form the background thought and subconscious force that drives much of the program. These are the Sato-Tate conjecture, the Ramanujan-Petersson conjecture, the Selberg eigenvalue conjecture, and Artin's conjecture about non-abelian L-functions attached to complex linear Galois representations. Then, we will give a brief and formal description of the construction of Langlands' L-functions and state the basic conjectures about them. In the final sections, we report on the progress made so far towards these conjectures highlighting those advances made in the last decade.

1. **The Sato-Tate conjecture.** Given an elliptic curve E over a number field F, we know that for each prime ideal v of F where E has good reduction, the number of points of E mod v is given by

$$N(v) + 1 - a_v$$

where $N(v)$ denotes the norm of v and a_v satisfies Hasse's inequality

$$|a_v| \leq 2\big(N(v)\big)^{1/2}.$$

One can therefore write

$$a_v = 2N(v)^{1/2} \cos \theta_v$$

where $0 \leq \theta_v \leq \pi$. The Sato-Tate conjecture is a statement about how the angles θ_v are distributed in the interval $[0, \pi]$ as v varies. In the case that E has

Received by the editors April 22, 2002.
Research partially supported by an NSERC grant.
AMS subject classification: 11-01, 11-02, 11Fxx, 11S37, 11S40.

complex multiplication, the behaviour is known and is a classical result of Hecke [H, pp. 215–234, 249–281]. In particular, Hecke's theorem says that if the CM field is contained in F, then the angles are uniformly distributed in $[0, \pi]$ in the sense of Weyl.

By contrast, in the non-CM case, Tate [T] and independently Sato conjectured that the angles are not uniformly distributed with respect to the usual Lebesgue measure but rather with respect to the measure

$$\frac{2}{\pi} \sin^2 \theta d\theta.$$

More precisely, they conjecture that the number of prime ideals v with norm less than x with $\theta_v \in (\alpha, \beta)$ is asymptotic as x tends to infinity, to

$$\left(\frac{2}{\pi} \int_\alpha^\beta \sin^2 \theta \, d\theta \right) \pi_F(x),$$

with $\pi_F(x)$ denoting the number of prime ideals with norm less than or equal to x. In a recent paper, Akiyama and Tanigawa [AT] make the stronger conjecture that the error term in this asymptotic formula is

$$O(x^{1/2+\epsilon})$$

for any $\epsilon > 0$. By applying Koksma's inequality in the theory of Diophantine approximation, they show that their conjecture implies that $L_1(s)$ (defined below) satisfies the analog of the Riemann hypothesis. A minor modification of their proof shows that in fact, all of the $L_m(s)$ (see below) satisfy the Riemann hypothesis as a consequence of their strong form of the Sato-Tate conjecture.

In the special case that E is defined over the rational number field, the modularity conjecture of Taniyama, Shimura and Weil (now a theorem) says that there exists a cusp form f_E of weight 2 (see definition in Section 2) such that for each prime p of good reduction for E, a_p is the p-th Fourier coefficient $a_p(f_E)$ of f_E. It then seems plausible that the Sato-Tate conjecture is really a statement about Fourier coefficients of cusp forms which are Hecke eigenforms. That this indeed should be the case, is predicted by Serre [Se]. More precisely, if f is a new form of weight k and level N, we may write

$$a_p(f) = 2p^{(k-1)/2} \cos \theta_p.$$

If f is not of CM-type, then Serre predicts that the angles θ_p are distributed with respect to the Sato-Tate measure

$$\frac{2}{\pi} \sin^2 \theta d\theta$$

as in the elliptic curve case.

In a fundamental paper, Langlands [La] outlined an approach to this conjecture via the theory of L-functions. We will be more precise about the origins of these L-functions in a later section. For now, they can be defined as follows.

Let $\alpha_p = e^{i\theta_p}$ and $\beta_p = e^{-i\theta_p}$. For each natural number m, set

$$L_m(s,f) = \prod_{(p,N)=1} \prod_{j=0}^{m} \left(1 - \frac{\alpha_p^{m-j}\beta_p^j}{p^s}\right)^{-1}.$$

These L-functions are referred to as symmetric power L-functions. Clearly, $L_m(s,f)$ converges absolutely for $\Re(s) > 1$. For $m = 0$, we have $L_0(s,f)$ is (apart from a few Euler factors) equal to the Riemann zeta function. Langlands [La] indicated that if each $L_m(s,f)$, $m \geq 1$, extends to an analytic function and does not vanish on the line $\Re(s) = 1$, then the Sato-Tate conjecture follows as a consequence. A more rigorous treatment and proof was later provided by Serre [Se]. Subsequently, Kumar Murty [VKM] showed that the non-vanishing hypothesis can be dropped since the analytic continuation and a superb generalization of the classical technique of Hadamard and de la Vallée Poussin establish that $L_m(1+it,f) \neq 0$. In this way, the Sato-Tate conjecture has been reduced to a problem of analytic continuation of certain L-functions.

Concerning the functions $L_m(s,f)$ for $m \geq 1$, Langlands conjectures that each of them extends to an entire function of s. This problem lies at the heart of the Langlands program and has been the muse of many discoveries and advances in the theory.

What is known about these analytic continuations? For $m = 1$, the assertion that $L_1(s,f)$ extends to an entire function is the work of Hecke [H]. It comes as a consequence of the theory of modular forms. Thus, encapsulated in $L_1(s,f)$ is a good chunk of what is called GL_2-theory which we will elaborate on below.

The method of Rankin [Ra] and Selberg [Sel] shows that

$$\zeta(s)L_2(s,f)$$

extends to $\Re(s) = 1$ except for a simple pole at $s = 1$. As the Riemann zeta function does not vanish on $\Re(s) = 1$, $s \neq 1$, and at $s = 1$, it has only a simple pole, we conclude that $L_2(s,f)$ extends to an analytic function for $\Re(s) \geq 1$. The Rankin-Selberg method however, does not allow us to conclude that $L_2(s,f)$ itself extends to an entire function since it does not exclude the possibility that $L_2(s,f)$ has poles at the zeros of $\zeta(s)$. Thus a new technique was needed to deduce the required property of $L_2(s,f)$. This is a famous theorem of Shimura [Sh] which signalled the famous Shimura correspondence between cusp forms of integral weight and certain cusp forms of half-integral weight.

Recently Shahidi and Kim [KS] have proved:

THEOREM 1. *$L_3(s,f)$ extends to an entire function.*

Very recently, Kim [K] has proved that

THEOREM 2. *$L_4(s, f)$ extends to an entire function.*

As a further consequence of their work, Kim and Shahidi [KS] have established that for $5 \le m \le 9$, $L_m(s)$ has meromorphic continuation to the entire complex plane.

These results represent tremendous progress in the Langlands program and have significant consequences to other questions we will consider, namely to the Ramanujan-Petersson conjecture and the Selberg eigenvalue conjecture which we will elucidate below.

2. Maass forms. Let $\mathcal{H}$ be the upper half plane. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on the upper half-plane via the fractional linear transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

It is convenient to consider the extended upper half-plane $\mathcal{H}^*$ which is $\mathcal{H}$ together with all the rational numbers and the point $i\infty$. These additional points (that is the set of rational numbers and $i\infty$) are called cusps. One extends the action of $\mathrm{SL}_2(\mathbb{Z})$ to $\mathcal{H}^*$ in the obvious way. Let $\Gamma_0(N)$ be the (Hecke) congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ defined by the set

$$\left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} (\operatorname{mod} N)\right\}.$$

Let χ be a Dirichlet character mod N. A modular form of weight k and level N and Nebentypus χ is defined as a holomorphic function $f\colon \mathcal{H}^* \to \mathbb{C}$ satisfying the transformation

$$f\left(\frac{az+b}{cz+d}\right) = \chi(d)(cz+d)^k f(z), \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

The set of all such forms comprises a finite dimensional $\mathbb{C}$-vector space, denoted $M_k(N,\chi)$ over the complex numbers. The subspace of forms vanishing at all the cusps is denoted $S_k(N,\chi)$. The elements of this space are called cusp forms. Since

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N),$$

the condition above implies that any modular form f has a Fourier expansion of the form

$$f(z) = \sum_{n=0}^{\infty} a_n(f) e^{2\pi i n z}.$$

Any such function $f(z)$ can also be viewed as a smooth function $f(x, y)$ of two real variables x, y and the holomorphy condition can be reformulated using the

Cauchy-Riemann equations. Consequently $f(x,y)$ is an eigenfunction of the non-Euclidean Laplacian

$$\Delta = -y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right).$$

If we drop the condition of holomorphy from the definition of a modular form and insist only that our function still is an eigenfunction of Δ, we arrive at the notion of a Maass wave form. Such forms necessarily have a Fourier expansion of the form

$$f(x,y)=\sum_{n=0}^{\infty} a_n(f,y)e^{2\pi inx}.$$

The condition that f be an eigenfunction of the Laplacian with eigenvalue λ means that f satisfies the (Whittaker) differential equation

$$-y^2\frac{d^2}{dy^2}a_n(f,y)=(\lambda-4\pi^2n^2y^2)a_n(f,y).$$

If we put $b_n(f,y) = a_n(f,y/2\pi n)/\sqrt{y/2\pi n}$, then $b_n(f,y)$ satisfies the famous Bessel equation:

$$y^2\frac{d^2}{dy^2}b_n(f,y)+y\frac{d}{dy}b_n(f,y)-\left(y^2+\left(\frac{1}{4}-\lambda\right)\right)b_n(f,y)=0.$$

The solutions of these equations have been well studied for over two centuries. For $n=0$, Whittaker's equation simplifies to

$$y^2\frac{d^2}{dy^2}a_0(f,y)=-\lambda a_0(f,y)$$

and it is easy to see that all solutions are of the form

$$b_0(y)=ay^s+by^{1-s},\quad \lambda=s(1-s).$$

Observe that if $\lambda=0$, that is, f is holomorphic, then the equation simplifies to

$$\frac{d^2}{dy^2}a_n(f,y)=4\pi^2n^2a_n(f,y)$$

and any solution is of the form

$$a_n(f,y)=a_n(f)e^{-2\pi ny}+a_n(f)'e^{2\pi ny}.$$

Since our functions are holomorphic at $i\infty$, the second solution drops out and we find that our solution for $\lambda=0$ has the form

$$f(x,y)=a_0+\sum_{n=1}^{\infty}a_n(f)e^{2\pi inz}$$

which is the familiar Fourier expansion of a modular form.

In the general case, when $\lambda \neq 0$, we write

$$\lambda = \frac{1}{4} + r^2$$

and derive the Fourier expansion for f as

$$f(z) = a_0(f)y^s + a_0'(f)y^{1-s} + \sum_{n \neq 0} a_n(f)\sqrt{y}K_{ir}(2\pi|n|y)e^{2\pi inx},$$

where

$$K_{ir}(y) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-y\cosh t}e^{-irt}\,dt.$$

Such a function is called a Maass wave form corresponding to the eigenvalue λ. If $f(z)$ is a Maass wave form, so is $\iota \circ f(z) = f(-\bar{z})$ with the same eigenvalue. As $\iota^2 = 1$, we can diagonalize the space of Maass forms with eigenvalue λ with respect to ι. If $\iota \circ f = f$, we say f is even and if $\iota \circ f = -f$, we say f is odd. Maass proved that if f is a cusp form so that $a_0(f)$ and $a_0(f)'$ both vanish, then the Dirichlet series

$$L(s,f) = \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s}$$

extends to an entire function and satisfies a functional equation of the form

$$\begin{aligned} &A^s\Gamma\left(\frac{s+\delta-ir}{2}\right)\Gamma\left(\frac{s+\delta+ir}{2}\right)L(s,f) \\ &\quad = wA^{1-s}\Gamma\left(\frac{1-s+\delta-ir}{2}\right)\Gamma\left(\frac{1-s+\delta+ir}{2}\right)L(1-s,\bar{f}) \end{aligned}$$

for a suitable constant A, and w of absolute value 1, and $\delta = 0$ or -1 depending on whether f is even or odd. (For further details about Maass forms, see [Bu, pp. 103–119].)

The Ramanujan-Petersson conjecture (still unresolved) predicts that

$$a_n(f) = O(n^\epsilon)$$

for any $\epsilon > 0$. The best result known as a consequence of the work of Shahidi and Kim is

$$a_n(f) = O(n^{7/64})$$

due to Kim and Sarnak [KSk].

Selberg's eigenvalue conjecture predicts that

$$\lambda \geq \frac{1}{4}.$$

Selberg himself proved it in the case $N = 1$ and showed in general that

$$\lambda \geq \frac{3}{16}.$$

The best result here, again due to Kim and Sarnak is

$$\lambda \geq \frac{975}{4096} = .238.$$

It is also an open problem to determine the dimension of the eigenspace corresponding to the eigenvalue λ. In the case $N = 1$, this space is conjectured to be one-dimensional.

All of these results are made possible because of our increasing knowledge of symmetric power L-functions. Indeed, in the paper alluded to above, Langlands [La] showed how one can construct for each Maass form, a family of L-functions $L_m(s, f)$ similar to the case of holomorphic f. However, unlike the holomorphic case, it is not known at present if all of these L-functions are analytic in a fixed half-plane. If this were so, Langlands showed that both the Ramanujan-Petersson and Selberg conjectures would follow. We will describe this in some more detail in the section below.

3. Artin's conjecture. As before, let F be an algebraic number field with ring of integers O_F. Let K/F be a finite Galois extension with group G. Given a prime ideal $\mathfrak{p}$ of F, we have by a classical theorem of Dedekind a unique factorization

$$\mathfrak{p}O_K = \wp_1^{e_1} \cdots \wp_g^{e_g}$$

into prime ideals $\wp_i$. For each prime ideal $\wp$ of K, we define the decomposition group

$$D_\wp = \{\sigma \in G : \sigma(\wp) = \wp\}$$

and a normal subgroup of $D_\wp$ called the inertia group

$$I_\wp = \{\sigma \in G : \sigma(x) \equiv x (\operatorname{mod} \wp), \forall x \in O_K\}.$$

The quotient group $D_\wp / I_\wp$ is canonically isomorphic to the Galois group of the finite field extension $O_K/\wp$ over $O_F/\mathfrak{p}$. It has a generator $\sigma_\wp$ called the Frobenius automorphism that is uniquely determined by the property

$$\sigma_\wp(x) \equiv x^{N_{F/\mathbb{Q}}(\mathfrak{p})} (\operatorname{mod} \wp).$$

Thus, $\sigma_\wp$ is determined uniquely modulo the inertia subgroup (which is trivial apart from a finite set of primes which are ramified in K). Its conjugacy class is denoted $\sigma_\mathfrak{p}$ and is called the Artin symbol of $\mathfrak{p}$ in the literature.

Now suppose that V is a finite dimensional representation of G. That is, we have a homomorphism

$$\rho: G \longrightarrow \mathrm{GL}(V),$$

with V a finite-dimensional vector space over $\mathbb{C}$. Artin attached an L-function to such a representation by setting:

$$L(s,\rho\,;K/F)=\prod_{\mathfrak{p}}\det\left(1-\rho(\sigma_{\mathfrak{p}})N_{F/\mathbb{Q}}(\mathfrak{p})^{-s}\mid V^{I_{\wp}}\right)^{-1}$$

where as remarked above $\sigma_{\wp}$ is uniquely determined modulo the inertia subgroup so that the determinant in the definition above is well-defined since it only depends on the conjugacy class of $\wp$ for any $\wp|\mathfrak{p}$ and we are restricting to the subspace fixed by the inertia subgroup $I_{\wp}$.

These L-functions have certain functorial properties. First, it is clear that they only depend on the character χ of ρ, so therefore, we denote $L(s,\rho\,;K/F)$ by $L(s,\chi\,;K/F)$. Also, if $\chi=\chi_1+\chi_2$, then

$$L(s,\chi\,;K/F)=L(s,\chi_1\,;K/F)L(s,\chi_2\,;K/F).$$

It is less clear that Artin L-series are invariant under the induction of characters. This was proved by Artin. Namely, if H is a subgroup of G and K^H is its fixed subfield, then for any character ψ of H, we have

$$L(s,\psi\,;K/K^H)=L(s,\mathrm{Ind}_H^G\,\psi\,;K/F).$$

Artin conjectured that for any irreducible character χ of G which is not trivial, $L(s,\chi\,;K/F)$ extends to an entire function. As a culmination of the development of class field theory, he showed that his conjecture is true for any one-dimensional character. This theorem embodies the famous reciprocity law of Artin [A].

Motivated by Artin's conjecture, Brauer [B] proved a fundamental theorem in group theory now called Brauer's induction theorem. It says that given any character χ of G we may write it as a $\mathbb{Z}$-linear combination

$$\chi=\sum_i n_i\,\mathrm{Ind}_{H_i}^G\,\psi_i$$

with ψ_i one-dimensional. Thus,

$$L(s,\chi\,;K/F)=\prod_i L(s,\mathrm{Ind}_{H_i}^G\,\psi_i\,;K/F)^{n_i}.$$

Now using the fact that Artin L-series are invariant under induction and the reciprocity law, we deduce that every Artin L-function extends to a meromorphic function of s.

One of the consequences of Artin's conjecture is Dedekind's conjecture that the Dedekind zeta function $\zeta_M(s)$ of a finite extension of F is always divisible by $\zeta_F(s)$. That is, $\zeta_M(s)/\zeta_F(s)$ is entire. If M/F is Galois, Dedekind's conjecture is known and this is the celebrated theorem of Aramata and Brauer. Dedekind's conjecture is also known to be true whenever M is contained in a solvable extension M of F. A simple proof can be found in [KM]. In the paper [KM] (see also [FM]), Kumar Murty proves by an exceedingly simple method the following result. Fix $s_0 \in \mathbb{C}$ and let n_χ be the order of the zero (or pole) $L(s, \chi, K/F)$ at $s = s_0$. Then

$$\sum_\chi n_\chi^2 \le r_0^2$$

where the sum is over all the irreducible characters of $\mathrm{Gal}(K/F)$ and r_0 is the order of the Dedekind zeta function $\zeta_K(s)$ at $s = s_0$. This implies the upper bound $|n_\chi| \le r_0$ and as a consequence, we can deduce that for any irreducible character χ,

$$L(s, \chi, K/F)\zeta_K(s)$$

is regular for all $s \neq 1$. It is actually possible to be a bit more precise. By a careful application of the Cauchy-Schwarz inequality in [KM], one can derive the sharper estimate $|n_\chi| < r_0$ if $n_\chi < 0$.

A primary goal of the Langlands program is to prove Artin's conjecture. In fact, certain L-functions already defined by Langlands have been shown to extend to entire functions by Godement and Jacquet [GJ]. The famous Langlands reciprocity conjecture is that each Artin L-function attached to an irreducible character is one of these functions already known to be entire.

This aspect of the program has had remarkable success. For instance, in the case of two dimensional Galois representations ρ, the following theorem is a consequence of the work of Langlands [La2] and Tunnell [Tu]:

THEOREM 3. *If ρ is a two-dimensional representation of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *with solvable image in* $\mathrm{PGL}_2(\mathbb{C})$, *then Artin's conjecture is true.*

In fact, more is true. The Artin L-function is automorphic (in the sense to be defined below). This theorem, can be viewed as a two-dimensional analogue of Artin's reciprocity law. This theorem, certainly being one of the pinnacles of 20th century mathematics was the starting point of the proof of Wiles [Wi] of the modularity conjecture referred to above. Combining this work with that of Ribet, one obtains a complete proof of the much celebrated Fermat's Last Theorem.

In case the image is non-solvable, some new results have been established, notably in the work [BDST].

4. The adele ring $\mathbb{A}_F$. To begin our description of the Langlands program, we need to introduce adeles and adele groups as well as discuss their representation theory. The notion of an adele can be motivated in the following way in the case of the rational number field. It can be regarded as an infinite tuple

$$a = (a_\infty, a_2, a_3, a_5, \dots)$$

with $a_\infty \in \mathbb{R}$ and $a_p \in \mathbb{Q}_p$, the p-adic number field, so that $a_p \in \mathbb{Z}_p$ (the ring of p-adic integers) for p sufficiently large. The set of all such adeles can be made into a ring $\mathbb{A}_\mathbb{Q}$ by componentwise addition and multiplication. Observe that the rationals can be viewed as a subring of the adeles via the embedding

$$x \mapsto (x, x, x, \ldots).$$

To each such a, we can define a character χ_a on the additive group of the rationals by setting

$$\chi_a(x) = \exp\bigl(2\pi i(-a_\infty x + a_2 x + a_3 x + \cdots)\bigr)$$

where we interpret the exponential as follows. For p sufficiently large, x is a p-adic integer and so $a_p x$ is a p-adic integer, so we ignore these in the evaluation of the exponential. For the remaining finitely many primes, we can write $a_p x$ as the sum of rational number (determined from the "polar part" of $a_p x$) and a p-adic integer, which again we ignore in the evaluation of the exponential. In this way, we view the argument of the exponential as a real number and evaluate the exponential as a function of a real variable. One can easily check that this gives a well-defined character of the additive group of the rationals. Moreover, two adeles a and b give rise to the same character if and only if

$$a - b = (x, x, x, \ldots)$$

for some rational number x. Thus, the character group of the rationals is isomorphic to $\mathbb{A}_\mathbb{Q}/\mathbb{Q}$. The adelic topology, defined below, is made so that this isomorphism is one of topological groups.

Generally, for an algebraic number field F, we define the adele ring $\mathbb{A}_F$ as follows. For each place v of F let F_v be the completion of F at v. Thus, in the case v is finite, F_v is the field of v-adic numbers and in the infinite case, it is either $\mathbb{R}$ or $\mathbb{C}$. As a set $\mathbb{A}_F$ consists of all infinite tuples (x_v) where $x_v \in F_v$ and for all but finitely many places, x_v lies in the ring O_v of v-adic integers. We make $\mathbb{A}_F$ into a ring by defining addition and multiplication componentwise. Then we impose the adelic topology on it by declaring that for each finite set S of places of F which contains the infinite places, the subring

$$\prod_{v \in S} F_v \times \prod_{v \notin S} O_v,$$

endowed with the product topology, is a basic open neighbourhood of the identity. One then checks that with this topology, $\mathbb{A}_F$ is a locally compact topological ring. Moreover, one can think of F as being embedded in $\mathbb{A}_F$ via the map

$$x \longmapsto (x, x, \ldots).$$

Now let $n \geq 1$ and $G = \mathrm{GL}_n$. For any commutative ring R with identity, $G(R)$ denotes the group of $n \times n$ matrices over R whose determinant is a unit of R. By

M_r we denote a generic $r \times r$ matrix and by I_r the $r \times r$ identity matrix. The so-called standard parabolic subgroups of GL_n can be described as follows. They are in one-to-one correspondence with the partitions $n = n_1 + \cdots + n_r$. For each such partition, the standard parabolic subgroup corresponding to this partition consists of matrices of the form

$$\begin{pmatrix} M_{n_1} & * & * \\ & \ddots & * \\ & & M_{n_r} \end{pmatrix},$$

and any parabolic subgroup is a GL_n conjugate of a standard parabolic subgroup. A parabolic subgroup has a Levi decomposition of the form $P = MN$ where N is the unipotent radical of P. M is called the Levi component of P. In the case of the standard parabolic, M and N can be described as consisting of matrices of the form

$$\begin{pmatrix} M_{n_1} & & \\ & \ddots & \\ & & M_{n_r} \end{pmatrix} \quad \begin{pmatrix} I_{n_1} & * & * \\ & \ddots & * \\ & & I_{n_r} \end{pmatrix}$$

respectively. We will write N_P to denote the unipotent radical in the Levi decomposition of a parabolic subgroup P.

The adelic topology on $G(\mathbb{A}_F)$ is similarly defined. Namely, for each finite set S as before, we declare

$$\prod_{v \in S} G(F_v) \times \prod_{v \notin S} G(O_v),$$

with the product topology, to be a basic open neighbourhood of the identity. The adelic topology on $G(\mathbb{A}_F)$ makes it into a locally compact group in which $G(F)$ is embedded diagonally as a discrete subgroup. If Z denotes the subgroup of scalar matrices in $G(\mathbb{A}_F)$, then the coset space $ZG(F) \setminus G(\mathbb{A}_F)$ with the quotient topology has finite volume with respect to any $G(\mathbb{A}_F)$-invariant measure. This coset space is to be viewed as the higher dimensional analogue of the upper half-plane discussed earlier. We will now consider certain square integrable functions on this space.

5. **Automorphic representations.** A character ω of $\mathbb{A}_F/F^*$ is called a Grossencharacter. We consider the Hilbert space

$$L^2(G(F) \setminus G(\mathbb{A}_F), \omega)$$

defined as the space of measurable functions ϕ on the coset space $G(F) \setminus G(\mathbb{A}_F)$ satisfying the conditions:

(i) $\phi(zg) = \omega(z)\phi(g)$ for all $z \in Z$ and $g \in G(F) \setminus G(\mathbb{A}_F)$;

(ii) $\int_{ZG(F)\setminus G(\mathbb{A}_F)} |\phi(g)|^2 \, dg < \infty$.

The subspace of cusp forms $L_0^2(G(F) \setminus G(\mathbb{A}_F), \omega)$ is defined by the extra condition

(iii) for all parabolic groups P of $G(\mathbb{A}_F)$,

$$\int_{N_P(F)\backslash N_P(\mathbb{A}_F)} \phi(ng)\, dn = 0$$

for every $g \in G(\mathbb{A}_F)$.

Let R be the right regular representation of $G(\mathbb{A}_F)$ on $L^2\big(G(F) \setminus G(\mathbb{A}_F), \omega\big)$ given by the action

$$\big(R(g)\phi\big)(x) = \phi(xg)$$

for $\phi \in L^2\big(G(F) \setminus G(\mathbb{A}_F), \omega\big)$ and $x, g \in G(\mathbb{A}_F)$. This is a unitary representation of $G(\mathbb{A}_F)$. An automorphic representation is a subquotient of the right regular representation of $G(\mathbb{A}_F)$ on $L^2\big(G(F) \setminus G(\mathbb{A}_F), \omega\big)$. A cuspidal automorphic representation is a subrepresentation of the right regular representation of $G(\mathbb{A}_F)$ on $L_0^2\big(G(F) \setminus G(\mathbb{A}_F), \omega\big)$.

A representation of $G(\mathbb{A}_F)$ is called admissible if its restriction to the maximal compact subgroup

$$K = \prod_{v \text{ complex}} U_n(\mathbb{C}) \times \prod_{v \text{ real}} O_n(\mathbb{R}) \times \prod_{v \text{ finite}} \mathrm{GL}_n(O_v)$$

contains each irreducible representation of K with only finite multiplicity.

To describe admissible representations, we need the notion of the restricted tensor product introduced in [JL]. Let $\{W_v : v \in V\}$ be a family of vector spaces indexed by V and V_0 a finite subset of V. In our case, V will be the set of places of F and V_0 the set of infinite places. For each $v \in V \setminus V_0$, let x_0 be a non-zero vector in W_v. For each finite subset S of V containing V_0, let

$$W_S = \bigotimes_{v \in S} W_v.$$

If $S \subset S'$, let

$$f_{S,S'}: W_S \longrightarrow W_{S'}$$

be defined by

$$\bigotimes_{v \in S} w_v \longmapsto \bigotimes_{v \in S} w_v \bigotimes_{v \in S' \setminus S} x_v.$$

Then, the restricted tensor product of the W_v with respect to the x_v, $W = \bigotimes_{x_v} W_v$ is defined as the inductive limit of the W_S. It is known (see [F, p. 181]) that any irreducible admissible representation of $\mathrm{GL}_n(\mathbb{A}_F)$ can be written as a restricted tensor product $\bigotimes_v \pi_v$ where π_v is an irreducible representation of $\mathrm{GL}_n(F_v)$. Moreover, the factors are unique up to equivalence.

For each finite v, one can construct certain representations of $\mathrm{GL}_n(F_v)$ as follows. Consider the Borel subgroup

$$B(F_v) = \left\{ \begin{pmatrix} b_1 & * & * \\ & \ddots & * \\ & & b_n \end{pmatrix} \right\} \subseteq \mathrm{GL}_n(F_v).$$

For any n-tuple of complex numbers, $z = (z_1, \ldots, z_n)$, define χ_z on $B(F_v)$ by

$$\chi_z(b) = |b_1|_v^{z_1} \cdots |b_n|_v^{z_n},$$

where $|\cdot|_v$ denotes the v-adic norm. This gives us a quasi-character on $B(F_v)$. Let $\tilde{\pi}_{v,z}$ be the representation of $\mathrm{GL}_n(F_v)$ obtained by inducing χ_z from $B(F_v)$ to $\mathrm{GL}_n(F_v)$. We shall assume

$$\Re(z_1) \geq \Re(z_2) \geq \cdots \geq \Re(z_n).$$

Then, the Langlands classification theorem (see [BW, Chapter 11, Section 2]) tells us that $\tilde{\pi}_{v,z}$ has a unique irreducible quotient $\pi_{v,z}$. The representations obtained in this way are called the unramified principal series. If π_v is any representation of $\mathrm{GL}_n(F_v)$ equivalent to $\pi_{v,z}$, we can define

$$A_v = \mathrm{diag}(Nv^{-z_1}, \ldots, Nv^{-z_n}) \in \mathrm{GL}_n(\mathbb{C})$$

where Nv denotes the norm of v.

Now let $\pi = \bigotimes_v \pi_v$ be an irreducible, admissible automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$. As π is admissible, there is a finite set S of places containing the infinite places, such that π_v belongs to the unramified principal series. For such v, define

$$L_v(s, \pi_v) = \det(1 - A_v Nv^{-s})^{-1}$$

and set

$$L_S(s, \pi) = \prod_{v \notin S} L_v(s, \pi_v).$$

It is possible to define $L_v(s, \pi_v)$ for $v \in S$ so that the complete L-function

$$L(s, \pi) = \prod_v L_v(s, \pi_v)$$

extends to an entire function and satisfies a functional equation when π is cuspidal. This is a theorem of Godement and Jacquet [GJ].

The Ramanujan conjecture in this full generality predicts that all the eigenvalues of A_v are of absolute value 1 for cuspidal representations. This is still one of the central unresolved problems in the Langlands program. Below, we will discuss what is known in the general case.

In the case $n = 1$, an automorphic representation is a Grossencharacter and L-functions of Grossencharacters were studied in detail by Hecke [H] who established their analytic continuations and functional equations. Hecke used classical methods and this was reworked using the adelic framework by Tate in his celebrated thesis [Ta].

In the case $n = 2$ and $F = \mathbb{Q}$, the L-function of a cuspidal automorphic representation is either the L-function of a holomorphic cusp form (attached to some weight and level) or the L-function of a Maass wave form. Deligne's work [D] establishes the Ramanujan conjecture for only the holomorphic cusp forms. Thus the case of Maass forms is still left open.

In the case $n = 3$, Hoffstein and Murty [HM] have given a more classical treatment of L-functions attached to automorphic representations.

There are several expository accounts of the Langlands program treating the general case that we may refer the reader to. The first one is Gelbart's article [G] but this is already 20 years old. Another one is due to Murty [M2]. A more recent report is Knapp's article [Kn] and a progress report of Henniart [He].

The famous reciprocity conjecture of Langlands asserts that every Artin L-function $L(s,\rho\ ;\ K/F)$ attached to an irreducible representation ρ of degree n must coincide with an $L(s,\pi)$ with π a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$. Since these latter L-functions have been shown to be entire, Artin's conjecture follows from the Langlands reciprocity conjecture.

As an extension of the modularity conjecture of Taniyama, Shimura and Weil, one expects that any reasonable L-function arising from arithmetic or a "motivic setting" should be the L-function of an automorphic representation. This is the general philosophy underlying the Langlands program.

6. Functoriality. In the general set up, one considers $G(\mathbb{A}_F)$ with G an arbitrary reductive algebraic group instead of GL_n and thus, one can define L-functions attached automorphic representations in a similar way. One expects that no new objects arise in this fashion and somehow everything is covered by GL_n. This is perhaps a rather diluted way of expressing the idea of functoriality. A more precise way is to say that whenever there is a homomorphism between two reductive groups, there should be a corresponding map between the spaces of cuspidal automorphic representations. This would imply an identity of the corresponding L-functions. We will expand on this theme in the special case of GL_n. In this, we will consider only two aspects of the functoriality conjecture.

Let us denote by $\mathcal{A}_n(F)$ to be the space of automorphic representations of $\mathrm{GL}_n(\mathbb{A}_F)$. For an arbitrary commutative ring R, we have the obvious map

$$\mathrm{GL}_n(R) \times \mathrm{GL}_m(R) \longrightarrow \mathrm{GL}_{nm}(R)$$

given by the tensor product of two matrices. This suggests that there should be a map

$$\mathcal{A}_n(F) \times \mathcal{A}_m(F) \longrightarrow \mathcal{A}_{nm}(F)$$

such that given two automorphic representations π and π', of degree n and m respectively, there should exist an automorphic representation Π in $\mathcal{A}_{nm}(F)$ so that

$$L(s,\Pi) = L(s,\pi \times \pi').$$

This conjecture makes two predictions. The first is the construction of an L-function $L(s, \pi \times \pi')$ from $L(s, \pi)$ and $L(s, \pi')$ and to establish its analytic continuation and functional equation. This has been done in a paper by Jacquet, Piatetskii-Shapiro and Shalika [JPS]. For $n = 2$, this is the classical work of Rankin [Ra] and Selberg [Sel2]. But the fact that it corresponds to an automorphic representation is conjectured by functoriality. For $n = 2$, this is a recent result of Ramakrishnan [Rm].

The symmetric power L-functions can be defined for any automorphic representation π in $\mathcal{A}_2(F)$. The simplest way to define them is as follows. As above, we have associated the matrix

$$A_v = \begin{pmatrix} \alpha_v & 0 \\ 0 & \beta_v \end{pmatrix}.$$

To this, we define the $(m+1) \times (m+1)$ matrix

$$A_v^{(m)} = \begin{pmatrix} \alpha_v^m & 0 & \cdots & 0 \\ 0 & \alpha_v^{m-1}\beta_v & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \beta_v^m \end{pmatrix}$$

and define the symmetric power L-function as

$$L(s, \mathrm{Sym}^m(\pi)) = \prod_v L_v(s, \mathrm{Sym}^m(\pi_v))$$

where for unramified v,

$$L_v(s, \mathrm{Sym}^m(\pi_v)) = \det(1 - A_v^{(m)} N v^{-s})^{-1}.$$

For the ramified places, a slight adjustment in the Euler factor is necessary. The main conjecture is that there should be an automorphic representation Π_m in $\mathcal{A}_{m+1}(F)$ such that

$$L(s, \Pi_m) = L(s, \mathrm{Sym}^m(\pi))$$

so that we have a hypothesized "lifting" of automorphic forms on GL(2) given by $\pi \mapsto \mathrm{Sym}^m(\pi)$. In [La], Langlands showed that if we could prove that each of the L-functions $L(s, \mathrm{Sym}^m(\pi))$ extends to an entire function and satisfies a functional equation, then the Sato-Tate conjecture, the Ramanujan conjecture and the Selberg eigenvalue conjecture follow immediately as a consequence. For the Sato-Tate conjecture, it is sufficient to have analytic continuation to $\Re(s) = 1$. For the Ramanujan and Selberg conjectures, the "weaker" assertion that they are analytic in some fixed half-plane is sufficent.

These L-functions, in the case that π corresponds to a holomorphic modular form coincide with the $L_m(s)$ defined in Section 1. In 1978, Gelbart and Jacquet [GJa] established that symmetric square lift exists. That is, they proved that $\mathrm{Sym}^2(\pi)$ is an automorphic representation of $\mathcal{A}_3(F)$, provided π is a non-monomial cuspidal representation. The recent work of Kim and Shahidi concerns the cases $m = 3$ and $m = 4$. It is now a theorem that the symmetric cube L-function is an entire function. For $m \leq 9$, we have meromorphic continuations of these functions. Below, we discuss the consequences of these results to the Selberg conjecture and the Ramanujan-Petersson conjecture. We summarise these results in:

THEOREM 4. *The following maps exist:*

$$\begin{array}{rll} \mathcal{A}_2(F) \longrightarrow \mathcal{A}_3(F), & \pi \longmapsto \mathrm{Sym}^2(\pi) & \textit{(Gelbart-Jacquet, 1978)} \\ \mathcal{A}_2(F) \longrightarrow \mathcal{A}_4(F), & \pi \longmapsto \mathrm{Sym}^3(\pi) & \textit{(Shahidi-Kim, 2000)} \\ \mathcal{A}_2(F) \longrightarrow \mathcal{A}_5(F), & \pi \longmapsto \mathrm{Sym}^4(\pi) & \textit{(Kim, 2001)} \\ \mathcal{A}_2(F) \times \mathcal{A}_2(F) \longrightarrow \mathcal{A}_4(F), & (\pi, \pi') \longmapsto \pi \times \pi' & \textit{(Ramakrishnan, 2000)} \\ & & \textit{(Kim, 2001)} \\ \mathcal{A}_2(F) \times \mathcal{A}_3(F) \longrightarrow \mathcal{A}_6(F), & (\pi, \pi') \longmapsto \pi \times \pi' & \textit{(Shahidi-Kim, 2000).} \end{array}$$

There is one more aspect of functoriality that is significant in the context of Artin's conjecture. This has to do with base change and automorphic induction. Very briefly, this can be explained as follows.

Given an automorphic representation Π in $\mathcal{A}_n(K)$ where K is a finite extension of F, we may consider the L-function $L(s, \Pi)$. If we group together the Euler factors lying above a given place v of F, we obtain an Euler factor that looks like one of the local Langlands L-functions attached to an automorphic representation of F_v. This, together with the fact that Artin L-functions are invariant under induction suggests that there is an automorphic representation $\pi \in \mathcal{A}_{nd}(F)$ where $d = [K : F]$ so that

$$L(s, \Pi) = L(s, \pi).$$

One refers to π as the automorphic induction of Π.

A dual notion is the concept of base change. Given any automorphic representation π of $\mathcal{A}_n(F)$, one may twist the representation by Dirichlet characters corresponding to an abelian extension K/F and thus get

$$\prod_{\chi} L(s, \pi \otimes \chi).$$

That there is an automorphic representation $\Pi \in \mathcal{A}_n(K)$ so that the above product equals $L(s, \Pi)$ is a famous theorem of Arthur and Clozel [AC]. In the case of a general Galois extension, one can make a similar construction but it is unknown at present whether Π exists. If it does, Murty [MRM] showed that Artin's conjecture follows immediately for all solvable extensions.

In fact, more can be proved. Suppose K/F is a finite extension of degree d and we have the automorphic induction map

$$\mathcal{A}_1(K) \longrightarrow \mathcal{A}_d(F).$$

Then both the Artin conjectures and Langlands reciprocity conjectures follow immediately.

To see this, we begin by writing an Artin L-function attached to an irreducible character χ of $\mathrm{Gal}(K/F)$ as

$$L(s,\chi\,;\,K/F)=\prod_i L(s,\mathrm{Ind}_{H_i}^G\,\psi_i\,;\,K/K^{H_i})^{n_i}$$

with ψ_i an abelian character of a proper subgroup H_i of G, as explained in Section 3. By the Artin reciprocity law, we obtain a Grossencharacter Π_i in $\mathcal{A}_1(K^{H_i})$ so that

$$L(s,\mathrm{Ind}_{H_i}^G\,\psi_i\,;\,K/K^{H_i})=L(s,\Pi_i).$$

By automorphic induction, we can write

$$L(s,\Pi_i)=L(s,\pi_i)$$

for $\pi_i\in\mathcal{A}_{d_i}(F)$ with $d_i=[K^{H_i}:F]$. After decomposing each π_i into cuspidal automorphic representations, we obtain

$$L(s,\chi,K/F)=\prod_j L(s,\pi_j)^{e_j}$$

with the π_j distinct (up to isomorphism) cuspidal automorphic representations and $e_j\in\mathbb{Z}$. Now consider the "Rankin-Selberg" L-function of both sides. By irreducibility and the fact that an Artin L-function is analytic at $s=1$ unless it corresponds to the trivial character (in which case it has a simple pole), we deduce that $L(s,\chi\otimes\overline{\chi},K/F)$ has a simple pole at $s=1$. On the other hand, the right hand side becomes a product of L-functions of the form $L(s,\pi_i\times\pi_j^*)$ which by the results [JPS] has no pole unless π_j^* equals the contragredient of π_i in which case it has a simple pole. We deduce

$$1=\sum_j e_j^2.$$

As the e_j's are integers, we deduce there is only one factor on the right hand side in the above factorisation $L(s,\pi_1)$ (say). Thus, $e_1=\pm1$. If $e_1=-1$, the Artin L-function will fail to have trivial zeros (and will have 'trivial poles') and so we conclude that $e_1=1$. This means

$$L(s,\chi,K/F)=L(s,\pi)$$

for some $\pi\in\mathcal{A}_n(F)$ with $n=\chi(1)$, which establishes both the Artin conjecture and Langlands reciprocity.

7. Converse theory. There seem to be few techniques available to establish the conjectures foreshadowed by the principle of functoriality. At present, there is the method of the trace formula, initiated by Selberg [Sel], and subsequently developed to arbitrary reductive algebraic groups by Langlands and Arthur. This method we shall not discuss here. The second method is described as converse theory and begins with a famous theorem of Weil [W].

The underlying idea is the following. Suppose we are given a sequence of complex numbers $\{a_n\}$ and we would like to show that there is a cusp form f of weight k and level N such that $a_f(n) = a_n$. If such a form exists, then clearly the Dirichlet series

$$L(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

extends to an entire function and satisfies a functional equation as implied by classical Hecke theory. In addition, the "twisting principle" tells us that if χ is a Dirichlet character $\bmod q$ with q coprime to N, then the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{a_n \chi(n)}{n^s}$$

also extends to an entire function and satisfies a functional equation which is precisely given in terms of Gauss sums, N, q and the Gamma function. The converse theorem of Weil says that these conditions are sufficient. A simplified proof of Weil's theorem was given later by Razar [Raz] who also reduced the infinite set of sufficient conditions of Weil to a finite set. Here is the precise form of Weil's theorem.

THEOREM 5 (WEIL). *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of complex numbers satisfying $a_n = O(n^c)$ for some constant $c > 0$. Fix a positive integer N and an even positive integer k. Let $\epsilon = \pm 1$. Suppose that*

$$\Lambda(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(s)$$

extends to an entire function, is bounded in vertical strips and satisfies a functional equation

$$\Lambda(s) = \epsilon(-1)^{k/2}\Lambda(k-s).$$

Suppose further that for every $(q, N) = 1$, and every primitive character χ $(\bmod q)$, the series

$$L_\chi(s) = \sum_{n=1}^{\infty} \frac{a_n \chi(n)}{n^s}$$

is such that

$$\Lambda_\chi(s) = (q^2 N)^{s/2}(2\pi)^{-s}\Gamma(s)L_\chi(s)$$

extends to an entire function, is bounded in vertical strips and satisfies

$$\Lambda_\chi(s) = \epsilon(-1)^{k/2}\frac{g_\chi}{g_{\overline{\chi}}}\chi(-N)\Lambda_{\overline{\chi}}(k-s)$$

where g_χ *is the Gauss sum*

$$g_\chi = \sum_{a \bmod q} \chi(a)e^{2\pi i a/q}.$$

Suppose further that $L(s)$ *converges absolutely at* $s = k - \delta$ *for some* $\delta > 0$*, then*

$$f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$$

is a cusp form of weight k *and level* N*.*

This theorem was generalized by Jacquet and Langlands [JL] to the context of automorphic forms on $\mathrm{GL}_2(\mathbb{A}_F)$ and recently to $\mathrm{GL}_n(\mathbb{A}_F)$ by Cogdell and Piatetskii-Shapiro [CPS].

Converse theory has been a main tool in several of the striking achievements of the Langlands program thus far. For example, the Gelbart-Jacquet lift [GJ], the work of Shahidi and Kim [KS] as well as the recent work of Ramkrishnan [Rm] establishing the modularity of the Rankin-Selberg L-function, all use converse theory in a fundamental way.

8. Estimating Fourier coefficients. An idea that goes back to Rankin [Ra] and Selberg [Sel2] is that if one has a sequence of complex numbers $\{a_n\}$ and the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{|a_n|^2}{n^s}$$

extends analytically to the entire complex plane apart from a simple pole at $s = 1$, and satisfies a functional equation, then, it is possible to estimate the coefficients a_n in a precise way. By methods of analytic number theory, one can derive a formula of the type

$$\sum_{n \le x} |a_n|^2 = rx + O(x^\theta)$$

with some non-zero constant r given by the residue at $s = 1$ and a value of $\theta < 1$ determined by the growth of the function (usually guided by the gamma factors in the functional equation). By putting $x = N$ and then $x = N - 1$ with N an integer and taking differences leads to the estimate $a_n = O(n^{\theta/2})$. It was this kind of method that first led to the significant advance on the Ramanujan-Petersson conjecture.

The technique actually goes back to Landau [L]. He proved:

THEOREM 6 (LANDAU). *Let $a_n \geq 0$ and set $f(s) = \sum_{n=1}^{\infty} a_n/n^s$. Suppose $f(s)$ is convergent in some half-plane and that it has analytic continuation for all $s \in \mathbb{C}$ except for a pole of order k at $s = 1$ and that it satisfies a functional equation*

$$Q^s\Delta(s)f(s) = Q^{1-s}\Delta(1-s)f(1-s)$$

where Q is a positive constant and

$$\Delta(s) = \prod_{i=1}^{N} \Gamma(\alpha_i s + \beta_i).$$

Set $d = 2\sum_{i=1}^{N} \alpha_i$. Then,

$$\sum_{n \leq x} a_n = xP_{k-1}(\log x) + O(x^{\frac{d-1}{d+1}} \log^{k-1} x),$$

where P_{k-1} is a polynomial of degree $k-1$.

An immediate corollary to this theorem is that $a_n = O(n^{\frac{d-1}{d+1}} \log^{k-1} n)$. Such a result when combined with the knowledge of analyticity of the various symmetric power L-functions leads to improved estimates that converge to the Ramanujan-Petersson conjecture and the Selberg eigenvalue conjecture. For instance, in the case $F = \mathbb{Q}$, and f is a Maass wave form, Kim and Shahidi [KSk] obtain $a_p(f) = O(p^{7/64})$. A similar estimate is derived for the Selberg conjecture. For the general case of an arbitrary number field F, one obtains an exponent of $1/9$ by the work of Shahidi.

This overview, by its brevity, omits discussion of other important results and contributors. For example, we did not discuss the triple product L-functions which are essential in the work of Ramakrishnan [Rm] and Kim [K]. We did not discuss the trace formula and the important work in this area. For these and other central features of the program, we direct the reader to the excellent expositions [Kn] and [He]. For the student who would like to have a more detailed instruction, we recommend [Bu] and [WL].

REFERENCES

[AC] J. Arthur and L. Clozel, *Simple algebras, base change and the advanced theory of the trace formula.* Ann. of Math. Stud. **120**, Princeton University Press, 1990.

[AT] S. Akiyama and Y. Tanigawa, *Calculation of values of L-functions associated to elliptic curves.* Math. Comp. (227) **68**(1999), 1201–1231.

[A] E. Artin, *Über eine neue Art von L-Reihen.* Abh. Math. Sem. Univ. Hamburg **3**(1923), 89–108; see also *Collected Papers.* Addison-Wesley, 1965, 105–124.)

[B] R. Brauer, *On Artin's L-series with general group characters.* Ann. of Math. (2) **48**(1947), 502–514.

[Bu] D. Bump, *Automorphic Forms and Representations.* Cambridge Stud. Adv. Math. **55**, Cambridge University Press, 1997.

[BDST] K. Buzzard, M. Dickinson, N. Shepherd-Baron and R. Taylor, *On icosahedral Artin representations.* Duke Math. J. (2) **109**(2001), 283–318.

[BW] A. Borel and N. Wallach, *Continuous cohomology, discrete sungroups and representations of reductive groups*. Ann. of Math. Stud. **94**, Princeton University Press, 1980.

[CPS] J. Cogdell and I. Piatetskii-Shapiro, *Converse theorems for* GL_n. Inst. Hautes Études Sci. Publ. Math. **79**(1994), 157–214.

[D] P. Deligne, *La conjecture de Weil, I*. Inst. Hautes Études Sci. Publ. Math. **43**(1974), 273–307.

[F] D. Flath, *Decomposition of representations into tensor products, automorphic forms, representations and L-functions*. Proc. Sympos. Pure Math. **33**(1979), Amer. Math. Soc., 179–184.

[FM] R. Foote and K. Murty, *Zeros and poles of Artin L-series*. Math. Proc. Camb. Phil. Soc. **105**(1989), 5–11.

[G] S. Gelbart, *An elementary introduction to the Langlands program*. Bull. Amer. Math. Soc. (2) **10**(1984), 177–219.

[GJa] R. Godement and H. Jacquet, *A relation between automorphic forms on* GL(2) *and* GL(3). Proc. Nat. Acad. Sci. U.S.A. **73**(1976), 3348–3350.

[GJ] R. Godement and H. Jacquet, *Zeta Functions of Simple Algebras*. Lecture Notes in Math. **260**, Springer, 1972.

[H] E. Hecke, *Mathematische Werke*. Springer-Verlag, Göttingen, 1959.

[He] G. Henniart, *Progrès récent en fonctorialité de Langlands*. Séminaire Bourbaki **890**(2000–2001).

[HM] J. Hoffstein and M. Ram Murty, *L-series of automorphic forms on* $\mathrm{GL}(3,\mathbb{R})$. In: Number Theory (eds. J.-M. de Koninck and C. Levesque), de Gruyter, 1989, 398-408.

[JPS] H. Jacquet, I. Piatetskii-Shapiro and J. Shalika, *Rankin-Selberg convolutions*. Amer. J. Math. **105**(1983), 367–464.

[Kn] A. Knapp, *Introduction to the Langlands program*. In: Representation Theory and Automorphic Forms (eds. T. N. Bailey and A. W. Knapp), Proc. Sympos. Pure Math. **61**(1997), 245–302.

[K] H. Kim, *Functoriality for the exterior square of* GL_4 *and symmetric fourth of* GL_2. To appear.

[KS] H. Kim and F. Shahidi, *Functorial Products for* $\mathrm{GL}_2 \times \mathrm{GL}_3$ *and the Symmetric Cube for* GL_2. Ann. of Math., to appear.

[KS2] ———, *Functorial products for* $\mathrm{GL}_2 \times \mathrm{GL}_3$ *and functorial symmetric cube for* GL_2. C. R. Acad. Sci. Paris Sér. I Math. (8) **331**(2000), 599–604.

[KSk] H. Kim and P. Sarnak, *Refined estimates towards the Ramanujan and Selberg conjectures*. Appendix to [K].

[JL] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2). Lecture Notes in Math. **114**, Springer-Verlag, Berlin-New York, 1974.

[L] E. Landau, *Über die anzahl der gitterpunkte in gewissen bereichen (Part 2)*. Göttinger Nachrichten, 1915, 209–243.

[La] R. P. Langlands, *Problems in the theory of automorphic forms*. In: Lectures in Modern Analysis and Applications, Lecture Notes in Math. **170**, Springer-Verlag, 1970, 18–86.

[La2] ———, *Base change for* GL(2). Ann. of Math. Stud. **96**, Princeton University Press, 1980.

[KM] V. Kumar Murty, *Holomorphy of Artin L-functions*. In: Proc. Ramanujan Centennial Conf., Ramanujan Math. Society, Chidambaram, 1987, 55–66.

[VKM] V. Kumar Murty, *On the Sato-Tate conjecture*. In: Number Theory Related to Fermat's Last Theorem (ed. N. Koblitz), Progr. Math. **26**, Birkhäuser, 1982.

[MRM] M. Ram Murty, *Selberg conjectures and Artin L-functions*. Bull. Amer. Math. Soc. (1) **31**(1994), 1–14.

[M2] ———, *A motivated introduction to the Langlands program*. In: Advances in Number Theory (eds. F. Gouvêa and N. Yui), Oxford University Press, 1993, 37–66.

[Rm] D. Ramakrishnan, *Modularity of the Rankin-Selberg L-series and multiplicity one for* SL(2), Ann. of Math. **152**(2000), 45–111.

[Ra] R. Rankin, *Contributions to the theory of Ramanujan's function* $\tau(n)$ *and similar arithmetic functions, I, II*. Proc. Cambridge, Phil. Soc. **35**(1939), 351–372; *III*. Proc. Cambridge Phil. Soc. **36**(1940), 150–151.

[Raz] M. Razar, *Modular forms on* $\Gamma_0(N)$ *and Dirichlet series*. Trans. Amer. Math. Soc.

231(1977), 489–495.

[Sel] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series.* J. Indian Math. Soc. **20**(1956), 47–81.

[Sel2] ———, *Bemerkung über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist.* Arch. Math. og Naturvidenskab B. **43**(1940), 47–50; see also *Collected Papers.* Springer-Verlag, 38–41.

[Se] J.-P. Serre, *Abelian ℓ-adic Representations and Elliptic Curves.* Benjamin, 1970.

[Sh] G. Shimura, *On the holomorphy of certain Dirichlet series.* Proc. London Math. Soc. (3) **31**(1975), 79–98.

[T] J. Tate, *Algebraic Cycles and Poles of Zeta Functions.* In: Arithmetic Algebraic Geometry (ed. F. G. Schilling), Harper and Row, New York, 1965.

[Ta] ———, *Fourier analysis in number fields and Hecke's zeta function.* Ph.D. Thesis, Princeton University, 1950; reprinted in Algebraic Number Theory (eds. J. Cassels and A. Fröhlich), Brighton, 1965, 305–347.

[Tu] J. Tunnell, *Artin's conjecture for representations of octahedral type.* Bull. Amer. Math. Soc. (New Series) (2) **5**(1981), 173–175.

[W] A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen.* Math. Ann. **168**(1967), 149–167.

[Wi] A. Wiles, *Modular elliptic curves and Fermat's last theorem.* Ann. of Math. (2) **141**(1995), 443–551.

[WL] W. C. Winnie Li, *Number Theory with Applications.* World Scientific, 1996.

Department of Mathematics and Statistics
Queen's University
Kingston, Ontario
K7L 3N6
email: murty@mast.queensu.ca

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 55–60

THE NEGATIVE PELL EQUATION

TOMMY BÜLOW

Presented by M. Ram Murty, FRSC

ABSTRACT. Criteria for the solvability of the negative Pell equation $x^2 - dp^2y^2 = -1$, d square-free, p prime, are proved by means of class field theory. The most important criterion deals with all d for which $\mathbb{Q}(\sqrt{-d})$ has an elementary abelian 2-class group and $p \equiv 5 \pmod 8$ or $p \equiv 9 \pmod{16}$.

RÉSUMÉ. À l'aide de la théorie du corps de classes on démontre des critères pour la resolubilité de l'équation négative de Pell $x^2 - dp^2y^2 = -1$, où d est un entier naturel sans carrés et p est un nombre premier. Le critère le plus important concerne le cas, où le 2-groupe des classes d'idéaux du corps $\mathbb{Q}(\sqrt{-d})$ est élémentaire et $p \equiv 5 \pmod 8$ ou $p \equiv 9 \pmod{16}$.

Introduction. The problem of deciding whether the negative Pell equation

$$x^2 - Dy^2 = -1, \quad D \in \mathbb{N} \text{ not square,} \tag{1}$$

has integer solutions is a classical problem in number theory which has not been completely solved. Many mathematicians have contributed to the problem. Fermat and Euler were two of the first to study the equation systematically. Dirichlet [1] proved (by elementary means) certain sufficient conditions for solvability (expressed in terms of the quadratic or biquadratic residue character of the prime factors of D). For more recent results see the bibliography.

Clearly, we can assume that $4 \nmid D$. Let $d > 1$ be a square-free integer and let $k > 1$ be a an odd integer. The equation (1) with D not square-free can be written

$$x^2 - dk^2y^2 = -1 \tag{2}$$

and the solvability of this equation is known to be equivalent to that of

$$x^2 - dk^2y^2 = -4. \tag{3}$$

From this we see that the problem in question is that of deciding whether the norm of the fundamental unit of the order of conductor k in $\mathbb{Q}(\sqrt{d})$ is 1 or -1. This has consequences for the structure of the corresponding ring class fields. When studying the existence of integral solutions to (2) one can, as is well-known, assume that $k = p$ is a prime number $\equiv 1 \pmod 4$. Of course, one can

Received by the editors April 27, 2001.
AMS subject classification: 11D09.

assume that (2) with $k = 1$ has a solution. We shall also assume that $p \nmid d$. It is not hard to show that if $\left(\frac{d}{p}\right) = -1$ and if $x^2 - dy^2 = -1$ is solvable, then $x^2 - dp^2y^2 = -1$ is solvable.

The remaining case, $\left(\frac{d}{p}\right) = 1$, is still not completely settled. Below we present some results concerning this case. These results can be proved by means of class field theory.

The criteria.

NOTATION. Let $d > 1$ be a square-free integer. Let $p_1, \ldots, p_r$ be the odd prime factors of d; let P_i be the prime ideal in $\mathbb{Q}(\sqrt{-d})$ above p_i. Let $\varepsilon = \frac{u+t\sqrt{d}}{2} > 1$ $(u, t \in \mathbb{Z})$ be the fundamental unit of $\mathbb{Q}(\sqrt{d})$. Let $x^2 - dy^2 = -1$ be solvable. Then $N(\varepsilon) = -1$ and it is readily verified that $\left(\frac{u}{p_i}\right) = 1$. If u is a biquadratic residue $\pmod{p_i}$, we say that p_i and P_i are of type I; otherwise, p_i and P_i are of type II. Let β be the number of p_i of type II.

Our most important result is the following:

THEOREM. *Let $d > 1$ be square-free and assume that $x^2 - dy^2 = -1$ has a solution. Let $p \equiv 1 \pmod 4$ be a prime number with $\left(\frac{d}{p}\right) = 1$; let $\mathfrak{p}$ be one of the two prime ideals in $\mathbb{Q}(\sqrt{-d})$ above p. Let the class number of $\mathbb{Q}(\sqrt{-d})$ be $h(\mathbb{Q}(\sqrt{-d})) = 2^z m$, $2 \nmid m$. Assume that $\mathfrak{p}^{2m}$ is a principal ideal.*

1. For $d \equiv 5 \pmod 8$: There is a relation

$$p^{m_0} p_1^{a_1} \cdots p_{r-1}^{a_{r-1}} = s^2 + dv^2; \quad a_1, \ldots, a_{r-1} \in \{0, 1\}; \quad m_0, s, v \in \mathbb{Z} \setminus \{0\};$$

with m_0 minimal (this implies $m_0 \mid m$). Put

$$\Sigma_1 := \sum_{p_i \text{ of type II}} a_i.$$

2. For $2 \mid d$: There is a relation

$$p^{m_0} p_1^{a_1} \cdots p_r^{a_r} = s^2 + dv^2; \quad a_1, \ldots, a_r \in \{0, 1\}; \quad m_0, s, v \in \mathbb{Z} \setminus \{0\};$$

with m_0 minimal (this implies $m_0 \mid m$). Put

$$\Sigma_2 := \sum_{p_i \text{ of type II}} a_i.$$

If $p \equiv 1 \pmod 8$, we write $p = a^2 + 16b^2$, $a, b \in \mathbb{Z}$.
Then for

A) $p \equiv 5 \pmod 8$:

1) $d \equiv 5 \pmod 8$:

$$x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Leftrightarrow 4 \mid sv.$$

2) $2 \mid d$:

$$x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Leftrightarrow 2 \nparallel v.$$

B) $p \equiv 9 \pmod{16}$:[1]

1a) $d \equiv 5 \pmod 8 \wedge 2 \mid \beta$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Leftrightarrow \\ \Big(2 \nmid b \wedge ((2 \mid \Sigma_1 \wedge 8 \mid sv) \vee (2 \nmid \Sigma_1 \wedge 4 \parallel sv))\Big) \\ \vee\Big(2 \mid b \wedge ((2 \mid \Sigma_1 \wedge 4 \parallel sv) \vee (2 \nmid \Sigma_1 \wedge 8 \mid sv))\Big). \end{gathered}$$

1b) $d \equiv 5 \pmod 8 \wedge 2 \nmid \beta$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Leftrightarrow \\ \Bigg(2 \nmid b \wedge \Big((2 \mid \Sigma_1 \wedge (4 \parallel s \vee 8 \mid v)) \vee (2 \nmid \Sigma_1 \wedge (8 \mid s \vee 4 \parallel v))\Big)\Bigg) \\ \vee\Bigg(2 \mid b \wedge \Big((2 \mid \Sigma_1 \wedge (8 \mid s \vee 4 \parallel v)) \vee (2 \nmid \Sigma_1 \wedge (4 \parallel s \vee 8 \mid v))\Big)\Bigg). \end{gathered}$$

2) $2 \mid d$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Leftrightarrow \\ \Big(2 \nmid b \wedge ((2 \mid \Sigma_2 \wedge 8 \mid v) \vee (2 \nmid \Sigma_2 \wedge 4 \parallel v))\Big) \\ \vee\Big(2 \mid b \wedge ((2 \mid \Sigma_2 \wedge 4 \parallel v) \vee (2 \nmid \Sigma_2 \wedge 8 \mid v))\Big). \end{gathered}$$

C) $p \equiv 1 \pmod{16}$:

1a) $d \equiv 5 \pmod 8 \wedge 2 \mid \beta$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Rightarrow \\ \Big(2 \mid b \wedge ((2 \mid \Sigma_1 \wedge 8 \mid sv) \vee (2 \nmid \Sigma_1 \wedge 4 \parallel sv))\Big) \\ \vee\Big(2 \nmid b \wedge ((2 \mid \Sigma_1 \wedge 4 \parallel sv) \vee (2 \nmid \Sigma_1 \wedge 8 \mid sv))\Big). \end{gathered}$$

1b) $d \equiv 5 \pmod 8 \wedge 2 \nmid \beta$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \textit{is solvable} \Rightarrow \\ \Bigg(2 \mid b \wedge \Big((2 \mid \Sigma_1 \wedge (4 \parallel s \vee 8 \mid v)) \vee (2 \nmid \Sigma_1 \wedge (8 \mid s \vee 4 \parallel v))\Big)\Bigg) \\ \vee\Bigg(2 \nmid b \wedge \Big((2 \mid \Sigma_1 \wedge (8 \mid s \vee 4 \parallel v)) \vee (2 \nmid \Sigma_1 \wedge (4 \parallel s \vee 8 \mid v))\Big)\Bigg). \end{gathered}$$

[1] In the following, the symbol $\wedge$ is the logical 'and'; the symbol $\vee$ is the logical 'or'.

2) $2 \mid d$:

$$\begin{gathered} x^2 - dp^2y^2 = -1 \quad \text{is solvable} \Rightarrow \\ \Big(2 \mid b \wedge \big((2 \mid \Sigma_2 \wedge 8 \mid v) \vee (2 \nmid \Sigma_2 \wedge 4 \parallel v)\big)\Big) \\ \vee \Big(2 \nmid b \wedge \big((2 \mid \Sigma_2 \wedge 4 \parallel v) \vee (2 \nmid \Sigma_2 \wedge 8 \mid v)\big)\Big). \end{gathered}$$

REMARK. If $x^2 - dy^2 = -1$ has a solution and the 2-sylow group of the class group of $\mathbb{Q}(\sqrt{-d})$ is elementary abelian, then the condition about $\mathfrak{p}^{2m}$ being principal is clearly fulfilled for all p and it is not hard to show that $d \equiv 5 \pmod 8$ or $2 \mid d$. We note that the theorem is a generalization of a result in [6] about the special case $d = 2$.

We shall briefly sketch the main ideas of the proof: Let $\varepsilon > 1$ be the fundamental unit of $\mathbb{Q}(\sqrt{d})$. In an elementary way one proves

LEMMA. *Assume that* $2^\lambda \parallel p - 1$. *Then*

$$x^2 - dp^2y^2 = -1 \quad \text{is solvable} \quad \Longleftrightarrow \quad \varepsilon^{\frac{p-1}{2^{\lambda-1}}} \equiv -1 \pmod p \quad (\text{in } \mathcal{O}_{\mathbb{Q}(\sqrt{d})}).$$

From now on, we consider only primes $p \equiv 9 \pmod{16}$ (the other cases are quite similar); also, let $d \neq 2$, *cf.* the above remark. In this case, the lemma and well-known theorems from algebraic number theory about the splitting of prime ideals imply that $x^2 - dp^2y^2 = -1$ is solvable if and only if $\mathfrak{p}$ splits totally in $\mathbb{Q}(\sqrt{\varepsilon}, i)$ and $\mathfrak{p}$ does not split totally in $\mathbb{Q}(\sqrt[4]{\varepsilon}, i)$ where $i = \sqrt{-1}$ (the fact that $p \equiv 1 \pmod 4$ ensures that we can adjoin i without problems).

The extension $\mathbb{Q}(\sqrt{\varepsilon}, i)/\mathbb{Q}(\sqrt{-d})$ is Galois with Galois group $\mathbb{Z}/4$; but the extension $\mathbb{Q}(\sqrt[4]{\varepsilon}, i)/\mathbb{Q}(\sqrt{-d})$ is not Galois for $d \neq 2$. The extension $\mathbb{Q}(\sqrt[4]{2\varepsilon}, i)/\mathbb{Q}(\sqrt{-d})$ is Galois with Galois group $\mathbb{Z}/8$. By well-known ramification theory we have:

$$\begin{gathered} \mathfrak{p} \text{ splits totally in } \mathbb{Q}(\sqrt[4]{\varepsilon}, i) \Leftrightarrow \\ \mathfrak{p} \text{ splits totally in } \mathbb{Q}(\sqrt{\varepsilon}, i) \wedge \\ \left(\left(\left(\frac{2}{p}\right)_4 = 1 \wedge \mathfrak{p} \text{ splits totally in } \mathbb{Q}(\sqrt[4]{2\varepsilon}, i)\right)\right. \\ \left.\vee\left(\left(\frac{2}{p}\right)_4 = -1 \wedge \mathfrak{p} \text{ does not split totally in } \mathbb{Q}(\sqrt[4]{2\varepsilon}, i)\right)\right). \end{gathered}$$

The solvability of our equation is, therefore, a question of the splitting of $\mathfrak{p}$ in *abelian* extensions of $\mathbb{Q}(\sqrt{-d})$; hence we can apply class field theory.

Let (in the sense of class field theory) H (resp. H_1) be the ideal group in $\mathbb{Q}(\sqrt{-d})$ corresponding to the abelian extension $\mathbb{Q}(\sqrt{\varepsilon}, i)$ (resp. $\mathbb{Q}(\sqrt[4]{2\varepsilon}, i)$). These groups have conductors dividing (8) resp. (4). Let $S_{\mathfrak{M}}$ denote the ray class group modulo the divisor $\mathfrak{M}$ in $\mathbb{Q}(\sqrt{-d})$. For simplicity, we assume from now on that d is even. One can show that the subgroup of principal ideals in H is

$$H \cap S_{(1)} = \{(1), (5)\} S_{(8)}.$$

Note that $(1+4\sqrt{-d}) \notin H$ but $(1+4\sqrt{-d})^2 \in H$. It is not hard to show that
i) $P_i \in H$ if P_i is of type I, and that
ii) $P_i \notin H$ but $P_i^2 \in H$ if P_i is of type II.

The existence of a relation $p^{m_0} p_1^{a_1} \cdots p_r^{a_r} = s^2 + dv^2$ can be proved by genus theory and if m_0 is minimal, it is not difficult to show that $m_0 \mid m$ and that, for a suitable sign of v,

$$\mathfrak{p}^{m_0} P_1^{a_1} \cdots P_r^{a_r} = (s + v\sqrt{-d}).$$

Hence

$$\begin{aligned} \mathfrak{p} \in H &\Leftrightarrow \mathfrak{p}^{m_0} \in H \\ &\Leftrightarrow \mathfrak{p}^{m_0} \cdot \prod_{P_i \text{ of type I}} P_i^{a_i} \cdot \prod_{P_i \text{ of type II}} \left(P_i(1+4\sqrt{-d})\right)^{a_i} \in H \\ &\Leftrightarrow \left(2 \mid \Sigma_2 \wedge (s+v\sqrt{-d}) \in H\right) \\ &\qquad \vee \left(2 \nmid \Sigma_2 \wedge (s+v\sqrt{-d})(1+4\sqrt{-d}) \in H\right) \\ &\Leftrightarrow (2 \mid \Sigma_2 \wedge 8 \mid v) \vee (2 \nmid \Sigma_2 \wedge 4 \parallel v). \end{aligned}$$

In a similar way, it is proved that $\mathfrak{p} \in H_1 \Leftrightarrow 4 \mid v$.

By combining the above statements, the criterion (for $p \equiv 9 \pmod{16}$ and even $d \neq 2$) follows. Note that $\left(\frac{2}{p}\right)_4 = 1$ is equivalent to $2 \mid b$ (if $p = a^2 + 16b^2$) by a classical result of Gauss.

EXAMPLE. Let $d = 10$. First note that $3^2 - 10 \cdot 1^2 = -1$. The class number of $\mathbb{Q}(\sqrt{-10})$ is $h = 2$, so the theorem can be used for all primes $\equiv 5 \pmod 8$ or $\equiv 9 \pmod{16}$. Consider the primes 41 and 809 which are $\equiv 9 \pmod{16}$ and for which 10 is a quadratic residue. As $41^1 \cdot 5^0 = 1^2 + 10 \cdot 2^2$,

$$x^2 - 10 \cdot 41^2 y^2 = -1 \quad \text{is not solvable.}$$

From $809^1 \cdot 5^0 = 13^2 + 10 \cdot 8^2$ and $809 = 5^2 + 16 \cdot 7^2$ it follows that

$$x^2 - 10 \cdot 809^2 y^2 = -1 \quad \text{is solvable.}$$

We also have two other results similar to the above theorem but because of lack of space we can not state them here.

The first result deals with $d \equiv 1 \pmod 8$ and $\mathfrak{p}^m$ a principal ideal (with notation as in the theorem).

The second result gives the existence of a criterion (of the above form) for an arbitrary *even* d and $p \equiv 5 \pmod 8$ or $p \equiv 9 \pmod{16}$.

References

1. P. G. L. Dirichlet, *Einige neue Sätze über unbestimmte Gleichungen*. Gesammelte Werke, Chelsea, New York, 219–236.
2. Y. Furuta, *Norm of Units of Quadratic Fields*. J. Math. Soc. Japan **2**(1959), 139–145.
3. Chr. U. Jensen, *On the Solvability of a Certain Class of non-Pellian Equations*. Math. Scand. **10**(1962), 71–84.
4. ———, *On the Diophantine Equation* $\xi^2 - 2m^2\eta^2 = -1$. Math. Scand. **11**(1962), 58–62.
5. P. Kaplan, *Sur le 2-groupe des classes d'idéaux des corps quadratique*. J. Reine Angew. Math. **283/284**(1976), 313-363.
6. J. Perrot, *Sur l'équation* $t^2 - Dy^2 = -1$. J. Reine Angew. Math. **102**(1888), 185–223.
7. L. Rédei, *Bedingtes Artinsymbol mit Anwendungen in der Klassenkörpertheorie*. Acta Math. Sci. Hung. **4**(1953), 1–29.
8. ———, *Die 2-Ringklassengruppe des quadratischen Zahlkörpers und die Theorie der Pellschen Gleichung*. Acta Math. Sci. Hung. **4**(1953), 31–87.
9. L. Rédei and H. Reichardt, *Die Anzahl der durch 4 teilbaren Invarianten der Klassengruppe eines beliebigen quadratischen Zahlkörbers*. J. Reine Angew. Math. **170**(1934), 69–74.
10. A. Scholz, *Über die Lösbarkeit der Gleichung* $t^2 - Du^2 = -4$. Math. Z. **39**(1935), 95–111.
11. F. Tano, *Sur quelques théorèmes de Dirichlet*. J. Reine Angew. Math. **105**(1889), 160–169.

Department of Mathematics
University of Copenhagen
DK-2100 Copenhagen
Denmark
email: tommy@math.ku.dk

C. R. Math. Rep. Acad. Sci. Canada Vol. 24 (2), 2002 pp. 61–66

A CHARACTERIZATION OF VERONESE IMBEDDINGS INTO COMPLEX PROJECTIVE SPACES BY CIRCLES

KAORU SUIZU, SADAHIRO MAEDA AND TOSHIAKI ADACHI

Presented by M. Ram Murty, FRSC

ABSTRACT. Veronese imbeddings of complex projective spaces into complex projective spaces are typical examples of Kähler immersions. In this paper we characterize these Kähler immersions by an extrinsic property of circles on the submanifolds.

RÉSUMÉ. Les plongements véronésiens d'un espace projectif complexe dans un espace projectif complexe sont des exemples de plongements kählériens. Nous montrons que ces plongements kählériens se caractérisent par une propriété extrinsèque des cercles sur les sous-variétés.

1. Introduction. We consider a Kähler isometric full immersion of a complex projective space $\mathbb{C}P^n(c)$ of constant holomorphic sectional curvature c into a complex projective space $\mathbb{C}P^N(\tilde{c})$ of constant holomorphic sectional curvature $\tilde{c}$. By virtue of the classification theorem ([C], [NO]) this Kähler immersion is nothing but a Kähler imbedding $f_k\colon \mathbb{C}P^n(c/k) \to \mathbb{C}P^N(c)$ given by

$$[z_i]_{0\leq i\leq n} \longmapsto \left[\sqrt{\frac{k!}{k_0!\cdots k_n!}}\, z_0^{k_0}\cdots z_n^{k_n}\right]_{k_0+\cdots+k_n=k},$$

where $[*]$ means the point of the projective space with the homogeneous coordinates $*$ and $N = (n+k)!/(n!\,k!) - 1$. We usually call f_k the k-th *Veronese imbedding.* The imbedding f_k has various geometric properties. For example, it is known that for each geodesic γ on the submanifold $\mathbb{C}P^n(c/k)$ the curve $f_k \circ \gamma$ lies on a k-dimensional real projective space $\mathbb{R}P^k(c/4)$ of curvature $c/4$ which is a totally real totally geodesic submanifold of the ambient space $\mathbb{C}P^N(c)$. By using this property J. S. Pak and K. Sakamoto characterize each Veronese imbedding (for details, see [PS]).

In this paper, for each Veronese imbedding f_k we study the curve $f_k \circ \gamma$ for a circle γ of curvature κ (≥ 0) on the submanifold $\mathbb{C}P^n(c/k)$. We then find that for each k and each κ the curve $f_k \circ \gamma$ is an integral curve of some Killing vector field in $\mathbb{C}P^N(c)$ (see Section 5). This implies that the curve $f_k \circ \gamma$ has constant

Received by the editors June 4, 2001.
AMS subject classification: Primary: 53B25; secondary: 53C40.
Key words and phrases: complex projective space, Veronese imbeddings, circle.

curvatures $\kappa_1, \ldots, \kappa_{d-1}$ along $f_k \circ \gamma$ in $\mathbb{C}P^N(c)$. Here, this positive integer d ($\geq k$) is the order of the curve $f_k \circ \gamma$ in the sense of Frenet formula. Paying particular attention to this geometric property, we shall provide a characterization of every Veronese imbedding, which is the main result in this paper (Theorem 1).

2. Circles and Kähler circles. A smooth curve $\gamma = \gamma(s)$ parametrized by its arclength s on a Riemannian manifold M is called a *circle* if there exist a field $Y = Y_s$ of unit vectors along γ and a constant κ (≥ 0) satisfying

$$\begin{cases} \nabla_{\dot\gamma}\dot\gamma = \kappa Y \\ \nabla_{\dot\gamma} Y = -\kappa\dot\gamma, \end{cases} \tag{2.1}$$

where $\dot\gamma$ denotes the unit tangent vector of γ and $\nabla_{\dot\gamma}$ the covariant differentiation along γ with respect to the Riemannian connection ∇ of M. The constant κ is called the *curvature* of the circle. A circle of null curvature is nothing but a geodesic. For each point $x \in M$, each orthonormal pair (u, v) of vectors at x and each positive constant κ, there exists locally a unique circle $\gamma = \gamma(s)$ on M with initial condition that $\gamma(0) = x$, $\dot\gamma(0) = u$ and $\nabla_{\dot\gamma}\dot\gamma(0) = \kappa v$.

For a circle γ on a Kähler manifold $(M, \langle\ ,\ \rangle, J)$ with complex structure J, we see from (2.1) that $\langle\dot\gamma, JY\rangle$ is constant along γ. Therefore it makes sense to define a *Kähler circle* as a circle γ satisfying the condition that $\dot\gamma$ and Y span a holomorphic plane, namely $Y = J\dot\gamma$ or $Y = -J\dot\gamma$. Note that if γ is a Kähler circle, then (2.1) reduces to

$$\nabla_{\dot\gamma}\dot\gamma = \kappa J\dot\gamma \quad \text{or} \quad \nabla_{\dot\gamma}\dot\gamma = -\kappa J\dot\gamma. \tag{2.2}$$

3. Main result. We denote by $M_N(c)$ a complex space form, which is locally complex analytically isometric to a complex projective space $\mathbb{C}P^N(c)$ when $c > 0$, a complex Euclidean space $\mathbb{C}^N$ when $c = 0$ or a complex hyperbolic space $\mathbb{C}H^N(c)$ when $c < 0$. The main purpose of this paper is to prove the following:

THEOREM 1. *Let $f: M_n \to M_N(c)$ be a Kähler isometric full immersion of an n (≥ 2)-dimensional Kähler manifold M_n into an N-dimensional complex space form $M_N(c)$ of constant holomorphic sectional curvature $c > 0$. Then the following conditions are equivalent.*

(1) For some positive integer k, the submanifold M_n is locally congruent to $\mathbb{C}P^n(c/k)$, $N = (n+k)!/(n!\,k!) - 1$ and f is locally equivalent to the k-th Veronese imbedding f_k.

(2) There exists $\kappa > 0$ satisfying that for each circle γ of curvature κ on the submanifold M_n the curve $f \circ \gamma$ has constant first curvature κ_1 along this curve.

In order to prove this theorem we recall some fundamental results. An isometric immersion $f: M \to \widetilde{M}$ is said to be *isotropic* at $x \in M$ if $\|\sigma(X, X)\|/\|X\|^2$ ($= \lambda(x)$) does not depend on the choice of $X (\neq 0) \in T_xM$, where σ is the second fundamental form of f. If the immersion is isotropic at every point, then the immersion is said to be isotropic. When the function $\lambda = \lambda(x)$ is constant on M, we call M a constant (λ-)isotropic submanifold. Note that a totally umbilic immersion is isotropic, but not *vice versa*. The following is well-known.

LEMMA A ([O]). *Let f be an isometric immersion of M into $(\widetilde{M}, \langle\ ,\ \rangle)$. Then f is isotropic at $x \in M$ if and only if the second fundamental form σ of f satisfies $\langle\sigma(u,u), \sigma(u,v)\rangle = 0$ for an arbitrary orthogonal pair $u, v \in T_xM$.*

For isometric Kähler immersions of a complex space form into another complex space form we have the following result which characterizes the Veronese imbedding f_k.

THEOREM B ([C], [NO]). *Let $f: M_n(c) \to M_N(\tilde{c})$ be a Kähler isometric immersion of a complex space form of constant holomorphic sectional curvature c into another complex space form of constant holomorphic sectional curvature $\tilde{c}$.*

(i) If $\tilde{c} > 0$ and f is full, then $\tilde{c} = kc$ and $N = (n+k)!/(n!\,k!) - 1$ for some positive integer k.

(ii) If $\tilde{c} \leq 0$, then $c = \tilde{c}$ and $M_n(c)$ is totally geodesic in $M_N(\tilde{c})$.

We are now in a position to prove Theorem 1.

(2) $\Rightarrow$ (1). Let $f: M_n \to M_N(c)$ be a Kähler isometric full immersion satisfying the condition (2). We shall show that the submanifold M_n is isotropic at its each point in the ambient manifold $M_N(c)$. We take a point $x \in M_n$ and choose an arbitrary orthonormal pair of vectors $u, v \in T_xM_n$. Let $\gamma = \gamma(s)$, $s \in I$ be a circle of curvature κ on the submanifold M_n with initial condition that $\gamma(0) = x$, $\dot\gamma(0) = u$ and $\nabla_{\dot\gamma}\dot\gamma(0) = \kappa v$ which is defined on an open interval I on $\mathbb{R}$. It follows from equation (2.1) and the formula of Gauss $\widetilde{\nabla}_X Z = \nabla_X Z + \sigma(X, Z)$ that

$$\widetilde{\nabla}_{\dot\gamma(s)}\dot\gamma(s) = \kappa Y_s + \sigma\big(\dot\gamma(s), \dot\gamma(s)\big), \quad s \in I. \tag{3.1}$$

Here $\widetilde{\nabla}$ is the Riemannian connection of the ambient space $M_N(c)$. By condition (2) the first curvature $\kappa_1 = \|\widetilde{\nabla}_{\dot\gamma}\dot\gamma\|$ of the curve $f \circ \gamma$ is constant, so that equation (3.1) implies $\|\sigma(\dot\gamma, \dot\gamma)\|$ is constant on I. Hence, denoting by D the connection of the normal bundle of M_n in $M_N(c)$, from (2.1) we obtain

$$\begin{aligned}
0 = \frac{d}{ds}\|\sigma(\dot\gamma,\dot\gamma)\|^2 &= 2\big\langle D_{\dot\gamma}\big(\sigma(\dot\gamma,\dot\gamma)\big), \sigma(\dot\gamma,\dot\gamma)\big\rangle \\
&= 2\langle(\bar\nabla_{\dot\gamma}\sigma)(\dot\gamma,\dot\gamma) + 2\sigma(\nabla_{\dot\gamma}\dot\gamma,\dot\gamma), \sigma(\dot\gamma,\dot\gamma)\rangle \\
&= 2\langle(\bar\nabla_{\dot\gamma}\sigma)(\dot\gamma,\dot\gamma), \sigma(\dot\gamma,\dot\gamma)\rangle + 4\kappa\langle\sigma(\dot\gamma,\dot\gamma), \sigma(\dot\gamma, Y)\rangle.
\end{aligned} \tag{3.2}$$

Here the covariant differentiation $\bar\nabla$ of the second fundamental form σ with respect to the connection in (tangent bundle)+(normal bundle) is defined as follows:

$$(\bar\nabla_X\sigma)(Y,Z) = D_X\big(\sigma(Y,Z)\big) - \sigma(\nabla_X Y, Z) - \sigma(Y, \nabla_X Z).$$

Evaluating equation (3.3) at $s = 0$, we get

$$(3.3) \qquad \langle (\bar{\nabla}_u \sigma)(u,u), \sigma(u,u)\rangle + 2\kappa\langle \sigma(u,u), \sigma(u,v)\rangle = 0.$$

On the other hand, for another circle $\rho = \rho(s)$ of the same curvature κ on the submanifold M_n with initial condition that $\rho(0) = x$, $\dot\rho(0) = u$ and $\nabla_{\dot\rho}\dot\rho(0) = -\kappa v$, we have

$$(3.4) \qquad \langle (\bar{\nabla}_u \sigma)(u,u), \sigma(u,u)\rangle - 2\kappa\langle \sigma(u,u), \sigma(u,v)\rangle = 0$$

which corresponds to equation (3.3). Thus, from (3.3) and (3.4) we can see that $\langle \sigma(u,u), \sigma(u,v)\rangle = 0$ for any orthonormal pair of vectors u, v at each point x of M_n, so that the submanifold M_n is isotropic in $M_N(c)$ by Lemma A. On the other hand we denote by R (resp. $\widetilde{R}$) the curvature tensor of M_n (resp. $M_N(c)$). We recall the Gauss equation

$$\langle \widetilde{R}(X,Y)Z, W\rangle = \langle R(X,Y)Z,W\rangle + \langle \sigma(X,Z), \sigma(Y,W)\rangle - \langle \sigma(X,W), \sigma(Y,Z)\rangle.$$

Since M_n is a Kähler submanifold in $M_N(c)$, from this equation and

$$\widetilde{R}(X,Y)Z = \frac{c}{4}(\langle Y,Z\rangle X - \langle X,Z\rangle X + \langle JY,Z\rangle JX - \langle JX,Z\rangle JY + 2\langle X,JY\rangle JZ),$$

we find that the holomorphic sectional curvature $K(X,JX)$ of M_n determined by a unit vector X is given by

$$K(X,JX) = \langle R(X,JX)JX, X\rangle = c - 2\|\sigma(X,X)\|^2.$$

In view of the above M_n being isotropic in $M_N(c)$ implies that the holomorphic sectional curvature $K(X,JX)$ is constant for every unit vector X at its each point x. Thus, due to the complex version of Schur's theorem we can see that our manifold M_n is a complex space form. Therefore from Theorem B we conclude that our submanifold M_n is locally congruent to a Veronese manifold.

(1) $\Rightarrow$ (2). For each Veronese imbedding $f_k\colon \mathbb{C}P^n(c/k) \to \mathbb{C}P^N(c)$ we see that $\|\sigma(X,X)\|^2 = c(k-1)/2k$ for any unit vector X at each point $x \in \mathbb{C}P^n(c/k)$ (see [Og]). Then we find that for an arbitrary circle γ of curvature κ on $\mathbb{C}P^n(c/k)$ the curve $f_k \circ \gamma$ has constant first curvature $\kappa_1 = \sqrt{\kappa^2 + \frac{c(k-1)}{2k}}$ in the ambient manifold $\mathbb{C}P^N(c)$.

4. **Characterization of complex parallel hypersurfaces.** As an immediate consequence of Theorem 1 and Theorem B we establish the following.

THEOREM 2. *Let M_n be an n (≥ 2)-dimensional complex hypersurface of a complex space form $M_{n+1}(c)$. Then there exists $\kappa > 0$ satisfying that every circle γ of curvature κ on the hypersurface M_n, considered as a curve in $M_{n+1}(c)$, has* *constant first curvature κ_1 along this curve if and only if M_n is totally geodesic in $M_{n+1}(c)$.*

Our aim here is to generalize this theorem. We pay particular attention to *Kähler circles* on a complex hypersurface. The following is a characterization of complex hypersurfaces with parallel second fundamental form in a complex space form.

THEOREM 3. *Let M_n be a complex hypersurface of a complex space form $M_{n+1}(c)$. Then the following conditions are equivalent.*

(1) There exists $\kappa \geq 0$ satisfying that every Kähler circle γ of curvature κ on the hypersurface M_n, considered as a curve in $M_{n+1}(c)$, has constant first curvature κ_1 along this curve.

(2) Either M_n is totally geodesic in $M_{n+1}(c)$, or $c > 0$ and M_n is locally congruent to a complex quadric $Q_n(\mathbb{C})$.

This is a generalization of the result of Ogiue and the second author, which corresponds to the case $\kappa = 0$ [MO].

PROOF OF THEOREM 3.

(1) $\Rightarrow$ (2). It follows from (2.2) and (3.3) that

$$\langle(\bar{\nabla}_u\sigma)(u,u), \sigma(u,u)\rangle + 2\kappa\langle\sigma(u,u), \sigma(u,Ju)\rangle = 0,$$

so that

$$\langle(\bar{\nabla}_u\sigma)(u,u), \sigma(u,u)\rangle = 0. \tag{4.5}$$

for every unit vector u at any point $x \in M_n$. By replacing u by Ju, we obtain

$$\big\langle(\bar{\nabla}_u\sigma)(u,u), J\big(\sigma(u,u)\big)\big\rangle = 0. \tag{4.6}$$

Since M_n is a complex hypersurface, we can see from (4.5) and (4.6) that $(\bar{\nabla}_u\sigma)(u,u) = 0$ for all u. This implies that the second fundamental form σ is parallel. Thus we find that the complex hypersurface M_n is totally geodesic in $M_{n+1}(c)$ or locally congruent to a complex quadric $Q_n(\mathbb{C})$ (*cf.* [NT]).

(2) $\Rightarrow$ (1). Note that our complex hypersurfaces M_n's have parallel second fundamental form in $M_{n+1}(c)$. By easy computation we know that every Kähler circle on M_n, consider as a curve in $M_{n+1}(c)$, has constant first curvature.

5. Remarks. We here make mention of other curvatures of curves in Theorems 1 and 3.

REMARK 1. It is known that in a rank one symmetric space M, every circle γ is an integral curve of some Killing vector field of M (for details, see [MO], [MT]). This, combined with the fact that every Veronese imbedding f_k is an equivariant mapping, yields that for each circle γ on $\mathbb{C}P^n(c/k)$ the curve $f_k \circ \gamma$ is an integral curve of some Killing vector field of the ambient space $\mathbb{C}P^N(c)$, so that all curvatures of $f_k \circ \gamma$ are constant.

REMARK 2. In case where M is a Hermitian symmetric space, every Kähler circle is an integral curve of some Killing vector field of M (see [AMU]). Hence the same reason as in Remark 1 shows that each Kähler circle on $Q_n(\mathbb{C})$, as a curve in $\mathbb{C}P^{n+1}(c)$, has constant curvatures.

REFERENCES

[AMU] T. Adachi, S. Maeda and S. Udagawa, *Simpleness and closedness of circles in compact Hermitian symmetric spaces.* Tsukuba J. Math. **24**(2000), 1–13.

[C] E. Calabi, *Isometric imbedding of complex manifolds.* Ann. Math. **269**(1982), 481–499.

[MO] S. Maeda and K. Ogiue, *Geometry of submanifolds in terms of behavior of geodesics.* Tokyo J. Math. **17**(1994), 347–354.

[MOh] S. Maeda and Y. Ohnita, *Helical geodesic immersions into complex space forms.* Geom. Dedicata **30**(1989), 93–114.

[MT] K. Mashimo and K. Tojo, *Circles in Riemmanian symmetric spaces.* Kodai Math. J. **22**(1999), 1–14.

[NO] H. Nakagawa and K. Ogiue, *Complex space forms immersed in complex space forms.* Trans. Amer. Math. Soc. **219**(1976), 289–297.

[NT] H. Nakagawa and R. Takagi, *On locally symmetric Kaehler submanifolds in a complex projective space.* J. Math. Soc. Japan **28**(1976), 638–667.

[O] B. O'Neill, *Isotropic and Kaehler immersions.* Canad. J. Math. **17**(1965), 905–915.

[Og] K. Ogiue, *Differential geometry of Kaehler submanifolds.* Adv. Math. **13**(1974), 73–114.

[PS] J. S. Pak and K. Sakamoto, *Submanifolds with d-planar geodesic immersed in complex projective spaces.* Tôhoku Math. J. **38**(1986), 297-311.

Department of Mathematics
Shimane University
Matsue, Shimane, 690-8504
Japan
email: suizu@math.shimane-u.ac.jp

Department of Mathematics
Shimane University
Matsue, Shimane, 690-8504
Japan
email: smaeda@math.shimane-u.ac.jp

Department of Mathematics
Nagoya Institute of Technology
Gokiso, Nagoya, 466-8555
Japan
email: adachi@math.kyy.nitech.ac.jp

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 67–71

NONVANISHING OF CERTAIN RANKIN-SELBERG L-FUNCTIONS

A. RAGHURAM

Presented by M. Ram Murty, FRSC

Abstract. In this article we prove that given a holomorphic cusp form f and any point s_0 in the complex plane, there is a holomorphic cusp form g such that the Rankin-Selberg L-function $L(s, f \times g)$ is non-zero at s_0.

Résumé. Dans cet article, on prouve le résultat suivant. Étant donné une forme holomorphe cuspidale f et un point quelquonque du plan complexe, il existe une forme holomorphe cuspidale g telle que la fonction $L(s, f \times g)$ de Rankin-Selberg n'est pas nulle à s_0.

The aim of this article is to prove that given a holomorphic cusp form f on the upper half plane $\mathfrak{h}$, given any point s_0 in the complex plane and given any positive integer l there is a holomorphic cusp form g of weight $l+1$, which is also an eigenform and a newform and such that the Rankin-Selberg L-function $L(s, f \times g)$ is non-zero at s_0.

One may try to prove such a theorem by averaging. Namely, by choosing a suitable set of 'possible g's' and taking the average of $L(s, f \times g)$ over this set and then isolating a dominant term and showing it is non-zero. In some sense the point of this paper is to say that once such an averaging has been done in one context [6] then some generalities from the theory of automorphic forms takes over and gives our nonvanishing theorem 'almost for free'. The main ingredients in our proof are the notion of base change and automorphic induction for automorphic representations of GL(2) (a general reference for which is [1]) and the main theorem of Rohrlich [6].

After the proof of the main theorem we make various remarks wherein we carefully analyze the choices we make in getting hold of the 'twist' g and in particular say what the level of g can be. For instance, if l is even, then it is possible to arrange the level to be a squarefree product of 2 primes relatively prime to N and one of these primes can be essentially arbitrary. We then point out variations of this theorem wherein either f and/or g can be a Maass cusp form. Without further ado we state and prove the main theorem.

Received by the editors June 20, 2001.
AMS subject classification: Primary: 11F66; secondary: 11F67, 11F70, 22E55.

THEOREM. *Let $f \in S_k(\Gamma_0(N), \psi)$, i.e., f is a holomorphic cusp form of weight k, level N, and nebentypus ψ. Let l be any positive integer. Let $s_0 \in \mathbb{C}$. Then there exists $g \in S_{l+1}(\Gamma_0(M), \xi)$ for some level M and some nebentypus ξ such that*

$$L(s_0, f \times g) \neq 0$$

where $L(s, f \times g)$ is the Rankin-Selberg L-function attached to f and g. We can take g to be a Hecke eigenform and also a newform.

PROOF. Let $\pi = \pi(f)$ be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$ associated to f. (See [3, Chapter 5].) Let K be an imaginary quadratic extension of $\mathbb{Q}$ with discriminant relatively prime to N, the level of f. Let $\Pi = \mathrm{BC}_{K/\mathbb{Q}}(\pi)$ be the base change of π to an automorphic representation Π of $\mathrm{GL}_2(\mathbb{A}_K)$. By the assumption on the discriminant of K we have that Π is also cuspidal. (See [5, Lemma 11.3].) Later on we will be refining our choice of K to have some control over the level of g.

Choose a grossencharacter χ_1 of $\mathbb{A}_K^\times / K^*$ whose infinity component, which is a character of $\mathbb{C}^*$, is given by

$$\chi_{1,\infty}(z) = \left(\frac{z}{|z|}\right)^l$$

and such that χ_1 is unramified at every finite unramified place of K. (It is possible to choose such a character; see Remark 2.) Let S denote a finite set of places of K which contains all the primes dividing Nd_K where d_K is the discriminant of K.

Consider $\Pi \otimes \chi_1$. Let s_1 be any complex number. Apply the main theorem of Rohrlich [6] to $\Pi \otimes \chi_1$ to get a grossencharacter χ_2 which is unramified on S, with trivial infinity component and such that

$$L(s_1, (\Pi \otimes \chi_1) \otimes \chi_2) \neq 0.$$

Let $\chi = \chi_1 \chi_2$. By the formalism of base change and automorphic induction we have

$$L(s, \mathrm{BC}_{K/\mathbb{Q}}(\pi) \otimes \chi) = L(s, \pi \times \mathrm{AI}_{K/\mathbb{Q}}(\chi)).$$

Let $\tau = \mathrm{AI}_{K/\mathbb{Q}}(\chi)$ be the automorphic induction of χ. By the choice of $\chi_{1,\infty}$ we have that χ is not Galois invariant (under the Galois group of K over $\mathbb{Q}$) which gives that τ is a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$.

With the choices made on χ_1 and χ_2 we claim that τ is the cuspidal automorphic representation $\pi(g)$ associated to some holomorphic cusp form g of weight $l+1$.

Now the infinity component τ_∞ of τ corresponds, via the local Langlands correspondence for $\mathrm{GL}_2(\mathbb{R})$, to $\mathrm{Ind}_{\mathbb{C}^*}^{W_\mathbb{R}}(\chi_\infty)$. (See [4].) Here $W_\mathbb{R}$ is the Weil group of $\mathbb{R}$. Since $\chi_\infty = \chi_{1,\infty}$ we get that τ_∞ is the discrete series representation D_{l+1} of $\mathrm{GL}_2(\mathbb{R})$. (See [4, equations (2.1b) and (3.4)]. As a word of warning, our D_{l+1}

is D_l in [4]. This is done to emphasize that a subscript n in the notation for the discrete series D_n should be the weight n of the modular form it comes from and is characterized in the representation by SO(2) acting on this form via the character $e^{i\theta} \mapsto e^{-in\theta}$.)

Let ω_τ denote the central character of τ. The infinity component of ω_τ, denoted $\omega_{\tau,\infty}$ is equal to ω_{τ_∞} the central character of τ_∞. Now ω_{τ_∞} is, via the local Langlands correspondence, the determinant of $\mathrm{Ind}_{\mathbb{C}^*}^{W_\mathbb{R}}(\chi_\infty)$. A pleasant exercise (whose details we omit) gives that $\omega_{\tau,\infty}$ is $(\mathrm{sgn})^{l+1}$ and so it is trivial on $\mathbb{R}_{>0}$.

Let $\mathfrak{f}_\tau$ be the conductor of τ and put $M = |\mathfrak{f}_\tau|$ and $\xi = \omega_\tau$. Apply Theorem 5.19 of [3] to get a $g \in S_{l+1}(\Gamma_0(M), \xi)$ such that $\tau = \pi(g)$. Note that there is an eigenform which is also a newform satisfying this. Hence what we have proved till now is that

$$L(s_1, \pi \times \tau) = L\big(s_1, \pi(f) \times \pi(g)\big) \neq 0.$$

Note that $L\big(s, \pi(f) \times \pi(g)\big) = L(s + t, f \times g)$ where t depends only on l and k. Now taking $s_1 = s_0 - t$ finishes the proof. ∎

REMARK 1 (CHOICE OF K, χ_1, AND χ_2). Let p be a prime not dividing N such that $p \equiv 3 \pmod 4$ and let $K = \mathbb{Q}(\sqrt{-p})$. So we have that p is the only prime which ramifies in K and let $p\mathcal{O}_K = \mathfrak{p}^2$. (Here $\mathcal{O}_K$ is the ring of integers of K.)

We show now that it is possible to choose a unitary character χ_1 of $\mathbb{A}_K^\times/K^*$ whose infinity component $\chi_{1,\infty}$ is the character which sends $z \in \mathbb{C}^*$ to $(z/|z|)^l$ and such that χ_1 is unramified at all finite places except possibly at $\mathfrak{p}$. (For any finite place v of K, we let K_v denote the completion of K at v and let U_v denote the corresponding group of units, and let U_v^1 denote the first filtration subgroup of U_v.)

Now $K^* \cap (K_\infty^* U_\mathfrak{p}^1 \prod_{v \neq \infty, \mathfrak{p}} U_v) = \{1\}$ and so we can extend $z \mapsto (z/|z|)^l$ to a character of $K^*(K_\infty^* U_\mathfrak{p}^1 \prod_{v \neq \infty, \mathfrak{p}} U_v)/K^*$ by making it trivial on all the finite factors. Also since $K^* \cap (K_\infty^* \prod_{v \neq \infty} U_v) = \{\pm 1\}$ we can further extend up to $K^*(K_\infty^* \prod_{v \neq \infty} U_v)/K^*$ by asking that the local character at $\mathfrak{p}$ take the value $(-1)^l$ on -1. By standard character theory of locally compact abelian groups this character can be extended to a character of $\mathbb{A}_K^\times/K^*$ since this group contains $K^*(K_\infty^* \prod_{v \neq \infty} U_v)/K^*$ as subgroup of finite index. *Note that the conductor of χ_1 is $\mathfrak{p}$ if l is odd and χ_1 is unramified at all finite places if l is even.*

In the proof of the main theorem, χ_2 was a unitary character of K coming out of Rohrlich [6]. It is possible to choose it a little carefully. The constraints on the conductor $\mathfrak{q}$ of χ_2 are dictated by [6, Proposition 2]. (See also [6, p. 394].) These are that the absolute norm $\mathbb{N}(\mathfrak{q})$ of $\mathfrak{q}$ should be a product of distinct primes, that $\mathbb{N}(\mathfrak{q}) \gg 0$ and the wide ray class number modulo $\mathfrak{q}$, denoted $h^*(\mathfrak{q})$, should satisfy $h^*(\mathfrak{q}) > \mathbb{N}(\mathfrak{q})^{1-\epsilon}$ for every $\epsilon > 0$. With this in mind, let $\mathfrak{q}$ be a prime in K such that $\mathfrak{q} \cap \mathbb{Q} = (q)$ splits in K and $\mathfrak{q}$ is not in S. It is an easy exercise to see that the wide ray class number satisfies $h^*(\mathfrak{q}) \geq \mathbb{N}(\mathfrak{q})/2$. Hence wide ray class characters of K modulo $\mathfrak{q}$ satisfy Rohrlich's requirement (as long as $\mathbb{N}(\mathfrak{q}) = q \gg 0$). *So χ_2 is a wide primitive ray class character of conductor a prime $\mathfrak{q}$ as above.*

REMARK 2 (THE LEVEL AND NEBENTYPUS OF g). The level M of g is the absolute value of $\mathfrak{f}_\tau$ the conductor of $\tau = \mathrm{AI}_{K/\mathbb{Q}}(\chi)$. The conductor of τ is the product of all its local conductors. We analyze this case by case. *At the prime p:* Here we need two further cases. *If l is even* then $\chi_{\mathfrak{p}}$ is unramified and the Langlands parameter corresponding to τ_p is $\mathrm{Ind}_{W_{K_\mathfrak{p}}}^{W_\mathbb{Q}}(\chi_\mathfrak{p})$. Using Proposition 4(b) on [2, p. 158] (which computes the local Artin conductor of an induced representation) gives that the conductor of τ_p is p. *If l is odd* then $\chi_\mathfrak{p}$ has level 1 and this same proposition gives that the conductor of τ_p is p^2. *At the prime q:* We have $q\mathcal{O}_K = \mathfrak{q}\mathfrak{q}'$ and $\chi_{\mathfrak{q}'}$ is unramified whereas $\chi_\mathfrak{q}$ has level 1. Hence τ_q is an irreducible principal series representation whose Langlands parameter is $\chi_\mathfrak{q} \oplus \chi_{\mathfrak{q}'}$ and this has conductor q. *At any prime other than p or q:* It is clear at that all these local representations are unramified.

To summarize what we have is that *the level M of g is p^2q if l is odd and is pq if l is even, where p and q are two primes such that $(N, pq) = 1$, $p \equiv 3 \pmod 4$, q splits in $K = \mathbb{Q}(\sqrt{-p})$ and also $q \gg 0$.*

Calculating the nebentypus ξ of g is a classical computation of Hecke and in 'modern' language it is the central character of τ and this is the determinant of the induced representation. For our set up this gives ξ, as a character of $\mathbb{A}_\mathbb{Q}^\times/\mathbb{Q}^*$, to be $\xi = \omega_{K/\mathbb{Q}}\chi$ where $\omega_{K/\mathbb{Q}}$ is the character associated to the quadratic extension $K/\mathbb{Q}$ and the χ on the right hand side is the restriction of χ to $\mathbb{A}_\mathbb{Q}^\times/\mathbb{Q}^*$.

REMARK 3 (A COUPLE OF VARIANTS OF THE THEOREM). One variant of the theorem is as follows. We can begin with f being a Maass cusp form, *i.e.*, f is a bounded, $\Gamma_0(N)$ invariant, real analytic function on the upper half plane $\mathfrak{h}$ and such that f is an eigenfunction for the Laplace-Beltrami operator on $\mathfrak{h}$. (This eigenfunction condition looks like $\Delta(f) = \frac{1-s^2}{4}f$ where s is a purely imaginary number or a real number with absolute value less than 1.) Now let $\pi = \pi(f)$ be the cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$ associated to f. The rest of the proof goes through *mutatis mutandis* to get hold of a holomorphic cusp form g with very similar kind of restrictions, such that $L(s_0, f \times g) \neq 0$.

Alternatively we begin with f being either a holomorphic cusp form or a Maass cusp form and then considering the base change of $\pi = \pi(f)$ to K, but now taking K to be a real quadratic extension. Since K is real quadratic the infinity part of $\mathbb{A}_K^\times$ is $\mathbb{R}^* \times \mathbb{R}^*$. So we can choose a unitary character χ_1 of $\mathbb{A}_K^\times/K^*$ such that $\mathrm{AI}_{K/\mathbb{Q}}(\chi_1)$ has a unitary principal series representation of $\mathrm{GL}_2(\mathbb{R})$ with central character trivial on $\mathbb{R}_{>0}$. Now going through the entire proof with this π and χ_1 we can get hold of a Maass cusp form g such that $L(s_0, f \times g) \neq 0$. (These are the Maass cusp forms as constructed in [3, Section 7C] but with a 'Rohrlich twist'.)

ACKNOWLEDGEMENTS. I thank Ram Murty who posed this question to me in an informal conversation. I thank Eknath Ghate, Kumar Murty and Joe Repka for some helpful comments.

REFERENCES

1. J. Arthur and L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*. Ann. of Math. Stud. **120**, Princeton University Press, 1990.
2. J. W. S. Cassels and A. Frohlich, *Algebraic Number Theory*. Academic Press, 1967.
3. S. Gelbart, *Automorphic forms on adele groups*. Ann. of Math. Stud. **83**, Princeton University Press, 1975.
4. A. Knapp, *Local Langlands correspondence: The Archimedean case*. Proc. Symp. Pure Math. **55**(1994), part II, 393–410.
5. R. Langlands, *Base change for* GL(2). Ann. of Math. Stud. **96**, Princeton University Press, 1980.
6. D. Rohrlich, *Nonvanishing of L-functions for* GL(2). Invent. Math. **97**(1989), 381–403.

Department of Mathematics
University of Toronto
100 St. George Street
Toronto, ON
M5S 3G3
email: raghuram@math.toronto.edu

Current Address:
School of Mathematics
Tata Institute of Fundamental Research
Dr. Homi Bhabha Road
Colaba, Mumbai 400005
India
email: raghuram@math.tifr.res.in

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 72–76

ON THE VANISHING OF CUBIC RECURRENCES

MANISHA KULKARNI AND B. SURY

Presented by M. Ram Murty, FRSC

ABSTRACT. We study cubic recurrences such that the associated polynomial has all roots real. We shall prove that such a recurrence can take the value 0 at the most three times. An example is the sequence $u_n = \sum_i (-1)^i \binom{n-2i}{i}$, where t is any positive real number less than 4/27. The recurrence here is $u_{n+3} = u_{n+2} - tu_n$.

RÉSUMÉ. Nous considérons une suite récurrente nonconstante de rang 3 dont le polynôme correspondant a toutes ses racines réelles. Nous montrons qu'une telle suite s'annule trois fois au plus. Nous montrons aussi que ce résultat est optimal et en donnons une application.

Introduction. Consider the sequence of integers $u_n = \sum_i (-1)^i \binom{n-2i}{i}$. It was proved in [B] that $u_3 = u_{12} = 0$ are the only times that u_n takes the value 0. The sequence u_n satisfies the recurrence relation $u_{n+3} = u_{n+2} - u_n$. To study the multiplicity of zeroes of a sequence defined by such recursions, two types of methods are in usage. One is a p-adic method employing a powerful theorem of Strassman (see [C, p. 62]), and the other method needs (as in the above paper) detailed information on the units of an algebraic number field associated to the recurrence relation in a natural way. In the above sequence, the polynomial associated is $f(X) = X^3 - X^2 - 1$. This has exactly one real root. Our interest is in recurrences such that the associated polynomial has all roots real. An example is the sequence $u_n = \sum_i (-1)^i \binom{n-2i}{i} t^i$ where t is any positive real number less than 4/27. The recurrence here is $u_{n+3} = u_{n+2} - tu_n$. We shall prove that a nonconstant cubic recurrence whose associated polynomial has only real roots, can vanish at the most thrice. This would, for instance, imply that for a given t with $0 < t < 4/27$, u_n can take the value 0 at the most three values of n. We employ rather elementary methods. The main result is the following:

THEOREM. *Let $u_{n+3} = au_{n+2} + bu_{n+1} + cu_n$ where a, b, c are real. Assume that the associated polynomial $f(X) = X^3 - aX^2 - bX - c$ has real, nonzero roots α_i, $i = 1, 2, 3$ such that $\alpha_i \neq \pm\alpha_j$ for $i \neq j$. Then, the sequence u_n takes the value 0 at the most for three values of n unless it is identically 0.*

Received by the editors June 10, 2001.
AMS subject classification: 11B37.

REMARKS. (I) It is important to note that there are nonconstant cubic recurrences which take a value (say, zero) infinitely often. For instance, consider the recurrence sequence

$$u_{n+3} = u_{n+2} - u_{n+1} + u_n$$

with the starting values $1, 0, 1$ takes exactly the values $1, 0, 1, 2$ repeated. Note that the corresponding polynomial $X^3 - X^2 + X - 1$ has the roots $1, \pm i$.

(II) There are also cubic recurrences which vanish only finitely many times but vanish more than thrice. For instance, it can be shown (using the p-adic method of Strassman alluded to above—see [C, p. 71]) that the cubic recurrence $u_{n+3} = 2u_{n+2} - 4u_{n+1} + 4u_n$ with the initial values $u_o = 0 = u_1$, $u_2 = 1$ vanishes for $n = 0, 1, 4, 6, 13, 52$. Of course, this has non-real roots also.

(III) The case of quadratic recurrences is similar but simpler. For instance, it was proved in [S] that one has the following identity

$$\sum (-1)^i \binom{n-i}{i} X^i (1+X)^{n-2i} = 1 + X + \cdots + X^n$$

in $\mathbf{Z}[X]$ and that this gives that $\sum(-1)^i\binom{n-i}{i}$ takes the values $(-1)^n$, $(-1)^{n-1}$ or 0 according as $n \equiv 0, 1$ or $2 \bmod 3$. Note that the recurrence satisfied by these numbers is $u_{n+2} = u_{n+1} - u_n$. The corresponding polynomial $X^2 - X + 1$ has complex roots. On the other hand, we shall point out at the end of the paper that any nonconstant sequence defined by a quadratic recurrence $u_{n+2} = au_{n+1} + bu_n$ where the polynomial $X^2 - aX - b$ has real roots α, β with $\alpha \neq \pm\beta$, will have only one zero at the most.

(IV) For any real t, the numbers $u_n = \sum_i (-1)^i \binom{n-ri}{i} t^i$ satisfy the recurrence $u_{n+r+1} = u_{n+r} - tu_n$. It seems plausible from our method of proof that one could address higher order recurrences too. However, we do not pursue it here.

Before starting the proof of the theorem, we show very briefly that for a recurrence as in the proposition with the roots α_i distinct, it is true that there are real numbers a_i such that, for any n, $u_n = \sum_{i=1}^3 a_i \alpha_i^n$. Although this is well known 'high school algebra', the authors could not point out a reference. Here is a quick proof.

Now $f(X) = X^3 - aX^2 - bX - c = \prod_{i=1}^3 (X - \alpha_i)$. Write $Z(t) = \sum_{n \geq 0} u_n t^n$, a formal power series. Multiplying the recurrence by t^{n+3} and summing over n, we get

$$Z(t) = \frac{(au_1 + bu_0 - u_2)t^2 + (au_0 - u_1)t - u_0}{at + bt^2 + ct^3 - 1}.$$

Note that the denominator is $t^3 f(1/t)$. If α_i are distinct, we would like to choose a_i such that $u_n = \sum_{i=1}^3 a_i \alpha_i^n$. This entails choosing a_1, a_2, a_3 such that

$$Z(t) = \frac{a_1}{1 - \alpha_1 t} + \frac{a_2}{1 - \alpha_2 t} + \frac{a_3}{1 - \alpha_3 t}.$$

REFERENCES

[B] L. Bernstein, *Zeros of the functions* $f(n) = \sum_{i=0}(-1)^i\binom{n-2i}{i}$. J. Number Theory 6(1974), 264–270.

[C] J. W. S. Cassels, *Local Fields*. Cambridge University Press, 1986.

[PS] G. Polya and G. Szego, *Problems and theorems in analysis*. Springer, New York-Heidelberg-Berlin, 1976.

[S] B. Sury, *A curious polynomial identity*. Nieuw Arch. Wisk. **11**(1993), 93–96.

Stat-Math Unit
Indian Statistical Institute
8th Mile Mysore Road
Bangalore 560 059
India
email: sury@isibang.ac.in

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 77–84

IMAGE DE-NOISING BY A MULTI-DIRECTIONAL DIFFUSION EQUATION

STEVEN J. DESJARDINS AND RÉMI VAILLANCOURT

Presented by Vlastimil Dlab, FRSC

ABSTRACT. **The diffusion equation, $u_t = u_{xx} + u_{yy}$, has been used successfully in the reduction of noise in digital images – the partial differential equation *smoothes* the noise, reducing its distortion of the image. But since the direct application of the PDE to an image may result in image details being smoothed over and lost, other schemes have been developed to control the rate and direction of the diffusion to smooth noise while preserving edges and details in the image. An alternative to these algorithms has been developed. A multi-directional diffusion equation is applied to the image in Fourier Space. This new algorithm is quick and simple and gives good results in that the effect of noise is reduced, while the details are preserved reasonably well.**

RÉSUMÉ. **L'équation de la diffusion, $u_t = u_{xx} + u_{yy}$, débruite bien une image: elle lisse le bruit et réduit la distorsion, mais peut lisser et faire perdre des détails. Des schémas plus complexes contrôlent le taux et la direction de la diffusion. Dans cette note on débruite une image au moyen d'une équation de la diffusion multidirectionnelle qu'on applique dans l'espace de Fourier. Le nouvel algorithme est simple et rapide; il débruite l'image et en préserve les détails.**

0.1. *Introduction.* The heat, or diffusion, equation in two dimensions,

$$u_t = u_{xx} + u_{yy},$$

has been used in the reduction of noise in digital images, the rationale being that noise represents pertubations to the image. Application of the diffusion equation will smooth over the pertubations. The problem with this approach is that edge data and fine details may be smoothed over as well, further distorting the image.

Several attempts have been made to correct this limitation [1]–[13] and to try to enhance edges by running diffusion backwards in time in the vicinity of an edge. Perona and Malik [11] were the first to try such an approach. Their idea was to replace the diffusion equation with an anisotropic diffusion equation,

$$u_t = \nabla \cdot \big(g(|\nabla u|)\nabla u\big),$$

Received by the editors September 26, 2001.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

AMS subject classification: 68U10, 65T50, 35K15, 35Q80.

Key words and phrases: image de-noising, multi-product filter, diffusion equation.

where $g(\cdot)$ is a non-negative, monotonically decreasing function with $g(0) = 1$. Diffusion is controlled by the function $g(\cdot)$. Along an edge, the gradient is large in magnitude and normal to it (since the edge is a contour). Diffusion is encouraged within regions (where ∇u is small), but not across the boundaries of regions (edges). So, $g(\cdot)$ is larger within regions and smaller at edges. The goal is to smooth in directions parallel to the edge, but not perpendicular to it to preserve the edge and to try to run the diffusion backwards perpendicular to the edge to enhance it.

Another group of researchers, including Alvarez, Lions and Morel [2]–[6], [9], have furthered this work and corrected limitations in Perona and Malik's scheme. This group discovered that Perona and Malik's scheme will actually enhance and not remove some types of noise and it is unstable, as the solutions to slightly different initial conditions may diverge. Also, Perona and Malik's scheme will require pre-filtering in the case of noisy images. They suggested some further extensions:

$$u_t = \nabla \cdot \big(g(|\nabla G_\sigma * u|)\nabla u\big),$$

where

$$G_\sigma(x) = C\sigma^{-1/2}\exp(-|x|^2/4\sigma)$$

is a Gaussian with variance σ and $*$ is convolution [9]. This model is like Perona and Malik's, with the function $g(\cdot)$ to control edge enhancement, but now there is a different argument that is a superior estimator. The need for pre-filtering noise is eliminated. A more improved scheme is:

$$u_t = g(|G * \nabla u|)|\nabla u|\nabla \cdot \frac{\nabla u}{|\nabla u|},$$

where G is a smoothing kernel (like a Gaussian) [5]. This last scheme corrects the drawbacks of the Perona and Malik scheme and the term

$$|\nabla u|\nabla \cdot \frac{\nabla u}{|\nabla u|}$$

ensures that diffusion proceeds in directions orthogonal to ∇u, not in the direction of ∇u. The term $g(|G * \nabla u|)$ controls the edge enhancement as in their previous scheme above. This group went on to formalize a set of axioms for image processing [2]–[4].

Torkamani-Azar and Tait [12] have suggested

$$u_t = \nabla \cdot \big(g(\nabla[h * u])\nabla u\big),$$

where

$$h(x,y) = \frac{\beta}{2}\exp(-\beta(|x| + |y|)),$$

and β is a constant. Their method was also developed to correct the limitations of Perona and Malik and to be simpler to implement when discretized. Better smoothing is achieved than in Perona and Malik [12]. Torkamani-Azar and Tait found that there was a trade-off between sharpening edges and removing noise when choosing the value of the constant β. Smaller β led to better noise removal, whereas larger β preserved edges better. It was thus desirable to run several iterations with small β for the first run and larger β for the rest to remove noise on the first pass and then enhance the edges after.

Torkamani-Azar and Tait give a convenient expression for the signal-to-noise ratio (SNR) that will be used [12]:

$$\text{SNR} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} u(i,j)^2}{\sum_{i=1}^{m}\sum_{j=1}^{n}[u(i,j) - U(i,j)]^2} = \frac{\|u\|_F^2}{\|u-U\|_F^2}, \tag{1}$$

where $[u(i,j)]$ and $[U(i,j)]$ represent the original and noisy images, respectively, as matrices and $\|\cdot\|_F$ is the Frobenius matrix norm. Ideally, if noise were perfectly removed from a noisy image, the result would be $u = U$ and SNR is infinite. In general, a higher SNR value signifies a better result, though visual observation is the true measurement, as two matrices may have the same norm and yet appear completely different when viewed as images.

The results from these schemes are good. Noise is significantly reduced and edges are preserved or enhanced. To implement these schemes, the partial differential equations (PDE's) have to be discretized into difference equations. The intricacies in the above schemes stem from the desire to distinguish edges from noise and preserve or enhance the edges while diffusing the noise away. The methods above attempt to control the direction of diffusion using the gradient of the image and then diffuse a little in some areas, diffuse more in others and run the diffusion backwards in time in other regions of the image.

It was thought that it might be possible to remove noise with PDE's in a much simpler way while preserving edges and details. An alternative approach is this: diffuse in all directions by a small amount, thereby reducing the distortion caused by the noise and yet not damaging the image details too much at the same time. And so, a multi-directional diffusion was attempted in Fourier Space. The results are good—the new algorithm is relatively quick (typical calculations take approximately 30 s on a Sun Ultra 5, running at 360 MHz with 256 MB RAM, with approximately two-thirds of the time required to generate the filter matrix), easy to use, simple, reduces some of the noise and manages to preserve details reasonably well.

0.2. *De-noising with the product filter.* To apply the diffusion equation in a specific direction, a change of variable is required. Diffusion in the x-direction is governed by $u_t = u_{xx}$, in the y-direction by $u_t = u_{yy}$. To diffuse in the direction of a line that makes an angle of θ with the x-axis, the governing equation would be:

$$u_t = \cos^2\theta u_{xx} + 2\sin\theta\cos\theta u_{xy} + \sin^2\theta u_{yy}.$$

The two-dimensional Fourier Transform of function $u(x, y)$ is:

$$\widehat{u}(\xi, \eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, y)e^{-i(\xi x+\eta y)}\, dx\, dy.$$

The Fourier Transform of the PDE would be:

$$\begin{aligned}\widehat{u}_t &= -\xi^2\cos^2\theta\widehat{u} - 2\xi\eta\sin\theta\cos\theta\widehat{u} - \eta^2\sin^2\theta\widehat{u}\\ &= -[\xi\cos\theta + \eta\sin\theta]^2\widehat{u}.\end{aligned}$$

The equation is easily solved in Fourier Space. The solution would then be

$$\widehat{u}(\xi, \eta, t) = \widehat{u}(\xi, \eta, 0)\exp(-[\xi\cos\theta + \eta\sin\theta]^2 t).$$

Also, application of the PDE to an image in Fourier Space would require multiplications, as compared to the finite differences required in xy-Space.

Now, if diffusion were to be applied in many directions, specified by angles θ_k, the PDE would be:

$$u_t = \sum_k[\cos^2\theta_k u_{xx} + 2\sin\theta_k\cos\theta_k u_{xy} + \sin^2\theta_k u_{yy}].$$

Or, in Fourier Space:

$$\widehat{u}_t = -\Big(\sum_k[\xi\cos\theta_k + \eta\sin\theta_k]^2\Big)\widehat{u},$$

which would have solution

$$\widehat{u}(\xi, \eta, t) = \widehat{u}(\xi, \eta, 0)\exp\Big(-\sum_k[\xi\cos\theta_k + \eta\sin\theta_k]^2 t\Big).$$

So, if $\widehat{u}(\xi, \eta, 0)$ is taken to be the Fourier Transform of a noisy image (*i.e.*, the noisy image is taken as the initial condition), then applying the multi-directional PDE in Fourier Space reduces to matrix multiplication with an appropriately chosen value of t.

The Discrete Fourier Transform of an image, which is represented by a matrix, say of size $m \times n$, will also be an image of the same dimensions, and hence also represented by an $m \times n$ matrix in MATLAB. MATLAB's `fftshift` function is used to move the DC to the centre of the matrix and so the origin of the $\xi\eta$ coordinate system in Fourier Space (corresponding to the matrix coordinate system of MATLAB) will be at the centre of the image matrix, more explicitly at position

$$((m + 1)/2, (n + 1)/2)$$

(which is the exact centre if both m and n are odd), with the ξ-axis running downwards and the η-axis to the right. And so, the (ξ, η) coordinates of matrix element (i, j) are

$$(\xi, \eta) = (i - (m + 1)/2, j - (n + 1)/2).$$

The algorithm to apply this multi-directional diffusion equation proceeds in the following manner. The number of directions, p, is chosen (typically 256). Then there will be p angles $\{\theta_k \mid 0 \leq k \leq p-1\}$, where

$$\theta_0 = 0, \theta_1 = \pi/p, \theta_2 = 2\pi/p, \ldots, \theta_{p-1} = (p-1)\pi/p.$$

For each θ_k, a matrix with elements

$$\left[\left(i - (m+1)/2\right)\cos\theta_k + \left(j - (n+1)/2\right)\sin\theta_k\right]^2$$

is generated. These matrices are summed to produce a matrix with elements

$$\sum_{k=0}^{p-1}\left[\left(i - (m+1)/2\right)\cos\theta_k + \left(j - (n+1)/2\right)\sin\theta_k\right]^2.$$

This matrix is multiplied by the chosen value of t (typically of the order of 10^{-4}), normalized by dividing by the number of directions, p, and multiplied by -1. This resulting matrix is exponentiated elementwise to produce the matrix with elements

$$\exp\Big(-\sum_{k=0}^{p-1}\left[\left(i - (m+1)/2\right)\cos\theta_k + \left(j - (n+1)/2\right)\sin\theta_k\right]^2 t/p\Big).$$

This final matrix is then the product filter matrix for the algorithm. The Fast Fourier Transform (using MATLAB's `fft2` function) of the image is multiplied elementwise by this filter matrix and the Inverse Fourier Transform (MATLAB's `ifft2`) is calculated to produce the smoothed (reduced-noise) image. Several experiments were run with different images to find optimal values for the filter parameters, t and p. It was found that $p = 256$ was large enough to include many directions, but small enough to keep the algorithm efficient. The optimal range for the t parameter was found to be 0.0001 to 0.0005 and so, $t = 0.0003$ was typically used in calculations.

0.3. *De-noising the cameraman image.* Figures 1 and 2 present results of the new product filter algorithm applied to the Cameraman image. In Figure 1, the product filter algorithm is applied to the image with random noise of the form $50\,\mathrm{rand}(m,n) - 25\,\mathrm{rand}(m,n)$, where $\mathrm{rand}(m,n)$ is MATLAB's random matrix generator which produces an $m \times n$ matrix with elements uniformly distributed in the interval $(0,1)$. The intensities in the original (pre-noise) image range from 0 to 255 (and are integers). It can be seen that with $t = 0.00001$, essentially nothing is done, whereas with $t = 0.01$ too much smoothing is allowed. Figure 2 presents the results of the new product filter algorithm, with $t = 0.0003$ and random noise of the form $50\,\mathrm{rand}(m,n)$, compared to MATLAB's *averaging* filter, which averages intensities over 3×3 neighbourhoods. Visually, the results are quite comparable, though the new product filter algorithm seems to do a slightly superior job with detail preservation (see the man's face and the camera). Also, the new algorithm gives a superior SNR value (24 versus 22 for MATLAB).

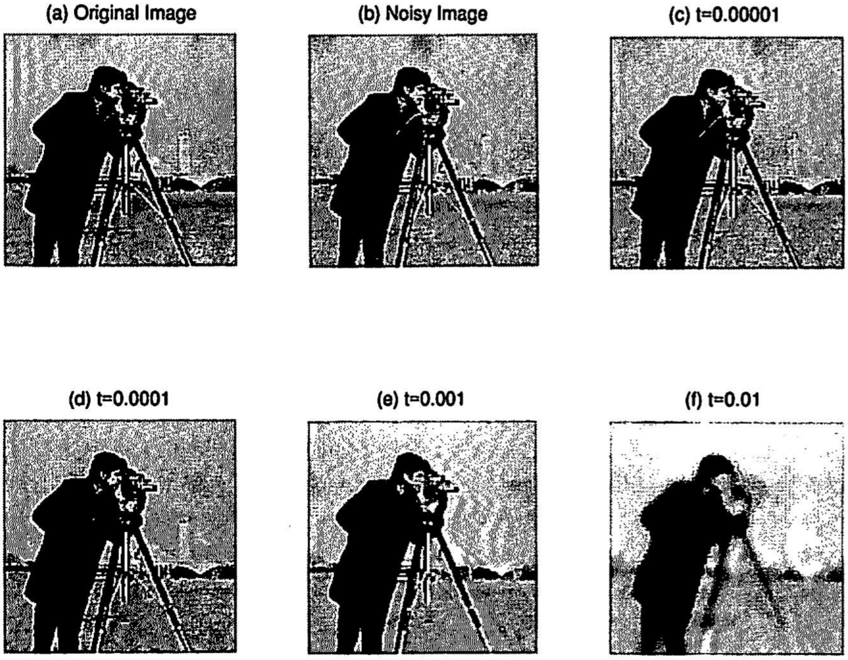

Figure 1: The Cameraman image with random noise, $50\,\mathrm{rand}(m,n) - 25\,\mathrm{rand}(m,n)$, for varying values of t and $p = 256$. (a) is the original (pre-noise) image. (b) is the image with the noise added. (c) to (f) are the processed images with the varying values of t shown above each sub-image. The SNR values are (b) 43.2793, (c) 46.3497, (d) 63.1046, (e) 45.1310 and (f) 22.4589.

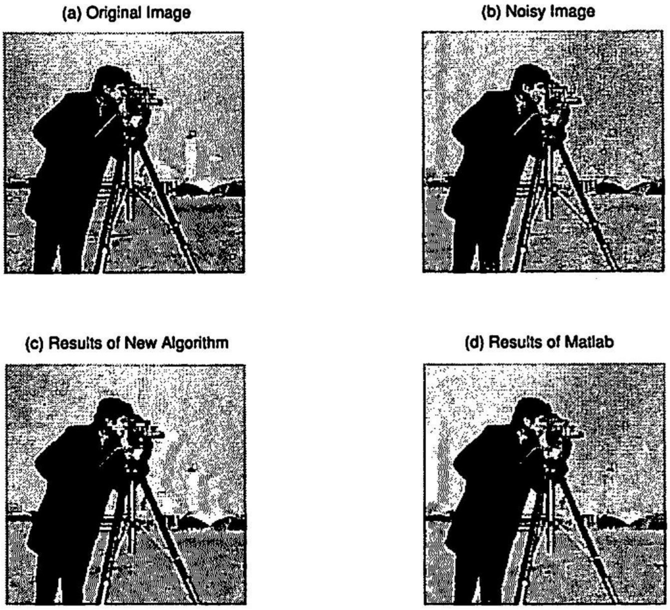

Figure 2: A comparison of the new product filter with MATLAB's *averaging* filter. $t = 0.0003$, $p = 256$ and the noise is $50\,\mathrm{rand}(m, n)$. (a) is the original image. (b) is the noisy image, SNR $= 21.7465$. (c) shows the results of the new product filter, SNR $= 24.3076$. (d) shows the results of MATLAB, SNR $= 22.1840$.

REFERENCES

1. S. T. Acton, *Multi-grid anisotropic diffusion.* IEEE Trans. Image Processsing (3) **7**(1998), 280–291.
2. L. Alvarez, F. Guichard, P.-L. Lions and J.-M. Morel, *Axiomes et équations fondamentales du traitement d'images (Analyse multiéchelle et E.D.P.).* C. R. Acad. Sci. Paris Sér. I **315**(1992), 135–138.
3. ———, *Axiomatisation et nouveaux opérateurs de la morphologie mathématique.* C. R. Acad. Sci. Paris Sér. I **315**(1992), 265–268.
4. ———, *Axioms and fundamental equations of image processing.* Arch. Rational Mech. Anal. **123**(1993), 199–257.
5. ———, *Image selective smoothing and edge detection by nonlinear diffusion II.* SIAM J. Numer. Anal. (3) **29**(1992), 845–866.
6. L. Alvarez and L. Mazorra, *Signal and image restoration using shock-filters and anisotropic diffusion.* SIAM J. Numer. Anal. (2) **31**(1994), 590–605.
7. M. J. Black, G. Sapiro, D. H. Marimont and D. Heeger, *Robust anisotropic diffusion.* IEEE Trans. Image Processsing (3) **7**(1998), 421–432.
8. V. Casselles, J.-M. Morel, G. Sapiro and A. Tannenbaum, *Introduction to the special issue on partial differential equations and geometry-driven diffusion in image processing.* IEEE Trans. Image Processsing (3) **7**(1998), 269–273.
9. F. Catté, P.-L. Lions, J.-M. Morel and T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion.* SIAM J. Numer. Anal. (1) **29**(1992), 182–193.
10. P. Perona, *Orientation diffusions.* IEEE Trans. Image Processsing (3) **7**(1998), 457–467.
11. P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion.* IEEE Trans. Pattern Anal. Mach. Intel. (7) **12**(1990), 629–639.
12. F. Torkamani-Azar and K. E. Tait, *Image recovery using the anisotropic diffusion equation.* IEEE Trans. Image Processing (11) **5**(1996), 1573–1578.
13. J. Weickert, B. M. ter Haar Romeny and M. A. Viergever, *Efficient and reliable schemes for nonlinear diffusion filtering.* IEEE Trans. Image Processsing (3) **7**(1998), 398–410.

Department of Mathematics and Statistics
University of Ottawa
Ottawa, Ontario
K1N 6N5
email: desjards@mathstat.uottawa.ca
remi@uottawa.ca

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 85–90

ON IRREDUCIBLE PARTIAL REPRESENTATIONS OF GROUPS

M. DOKUCHAEV AND N. ZHUKAVETS

Presented by Vlastimil Dlab, FRSC

ABSTRACT. We prove that there is a one-to-one correspondence between the irreducible finite degree partial representations of a group G over a field K and the (usual) irreducible K-representations of certain groupoid algebras of connected groupoids constructed from G. We give a structural result about irreducible partial representations on finite dimensional vector spaces over K and present the description "up to (usual) representations" of irreducible partial representations of abelian groups of degrees ≤ 3.

RÉSUMÉ. Nous démontrons qu'il y a une correspondence biunivoque entre les représentations partielles irréductibles de degré fini sur le corps K du group G et les K-représentations (usuelles) irréductibles de certaines algèbres de groupoïdes de groupoïdes connexes construites en partant de G. Nous donnons un résultat structurel sur les représentations partielles irréductibles en espaces vectorielles de dimension finie sur K et nous présentons la description, "à moins de représentations (usuelles)" des représentations partielles irréductibles des groupes abéliennes de degré ≤ 3.

Partial representations of groups appeared in the theory of operator algebras as an effective tool of their study (see [3] and [6]). In particular, the generators of the Cuntz-Krieger algebras [1] are images of the free generators of a free group (of finite rank) with respect to a partial representation (see [4], [5]), and this fact leads to a deeper understanding of their structure.

By a *partial representation* of a group G on a vector space V over a field K we mean a map

$$\pi: G \longrightarrow \mathrm{End}(V),$$

which sends the unit group element to the identity operator on V, such that for all $g, h \in G$ we have

$$\pi(g)\pi(h)\pi(h^{-1}) = \pi(gh)\pi(h^{-1}) \quad \text{and} \quad \pi(g^{-1})\pi(g)\pi(h) = \pi(g^{-1})\pi(gh).$$

Every K-representation of G on V obviously satisfies this definition but the partial representations are far from being exhausted by the representations in

Received by the editors December 5, 2001.

The first author was supported partially by CNPq of Brazil (Proc. 301115/95-8) and partially by Fapesp of Brazil (Proc. 01/05305-7).

AMS subject classification: Primary: 16S10; Secondary: 20C07.

Key words and phrases: partial representations of groups, groupoid algebras.

the usual sense of the word. If we replace $\mathrm{End}(V)$ by an arbitrary unital K-algebra $\mathcal{A}$, we get a more general notion of partial representation. In particular, if $\mathcal{A} = \mathrm{End}(W_R)$, where W is a free right module over a ring R, then π becomes a partial R-representation of G on W.

The group algebra KG is responsible for the theory of representations of G on K-vector spaces; similarly, there is an associative algebra $K_{\mathrm{par}}(G)$, called the partial group algebra of G, which governs the partial representations of G. The *partial group algebra* $K_{\mathrm{par}}(G)$ can be defined as the semigroup algebra $KS(G)$ of the semigroup $S(G)$, generated by the symbols $\{[g] : g \in G\}$ with relations $[g^{-1}][g][h] = [g^{-1}][gh]$, $[g][h][h^{-1}] = [gh][h^{-1}]$, $g, h \in G$, and $[e] = 1$, where e and 1 are the identity elements of G and $S(G)$, respectively. It is shown in [3] that $S(G)$ is an inverse semigroup. In [2] the structure of the partial group algebras $K_{\mathrm{par}}(G)$ has been determined for a finite group G and the isomorphism problem for these algebras was studied. The structure of $K_{\mathrm{par}}(G)$ was described by constructing for a finite group G a groupoid $\Gamma(G)$ whose groupoid algebra $K\Gamma(G)$ is isomorphic to $K_{\mathrm{par}}(G)$ and has the same representation theory as the partial representation theory of G.

Given a partial K-representation $\pi: G \to \mathrm{End}(V)$, V can be considered as a *partial G-space*, that is, a vector space over K with a product $G \times V \to V$ satisfying the conditions:

(a) $ex = x$;

(b) $g^{-1}\big(g(tx)\big) = g^{-1}\big((gt)x\big)$;

(c) $g\big(t(t^{-1}x)\big) = (gt)(t^{-1}x)$,

for all $g, t \in G$ and $x \in V$.

If V is finite dimensional, taking a basis in V we obtain the corresponding partial matrix representation $\pi: G \to M_n(K)$. Two partial matrix representations $\pi_1: G \to M_n(K)$ and $\pi_2: G \to M_n(K)$ are *equivalent* if there exists an invertible matrix $C \in \mathrm{GL}_n(K)$ such that $C^{-1}\pi_1(g)C = \pi_2(g)$ for all $g \in G$. It follows that two partial representations $\pi_i: G \to \mathrm{End}(V_i)$, $i \in \{1,2\}$, are *equivalent* if there exists a K-vector space isomorphism $\varphi: V_1 \to V_2$ such that $\varphi\pi_1(g) = \pi_2(g)\varphi$ for all $g \in G$. We say that $\pi: G \to \mathrm{End}(V)$ is *reducible*, if V contains an invariant partial G-subspace $V' \subsetneq V$. Otherwise, V is called *irreducible.*

By a *groupoid* Γ we mean a small category in which every morphism is an isomorphism. We may represent a groupoid as an oriented graph whose vertices are the objects and the oriented edges are the morphisms. We identify Γ with its set of morphisms. Every morphism $\gamma \in \Gamma$ has a source (domain) $s(\gamma)$ and a range (codomain) $r(\gamma)$ which are the corresponding objects of the category. A connected component of Γ is a maximal subgroupoid connected by morphisms. Then the *groupoid algebra* $K\Gamma$ is a K-vector space whose basis is Γ and with the multiplication given by

$$\gamma_1 \cdot \gamma_2 = \begin{cases} \gamma_1\gamma_2, & \text{if the composite morphism } \gamma_1\gamma_2 \text{ exists in } \Gamma, \\ 0, & \text{otherwise.} \end{cases}$$

If the number of objects of Γ is finite then the unity of $K\Gamma$ is the sum of all identity morphisms.

Given an arbitrary group G the groupoid $\Gamma(G)$ is defined as follows. The elements of $\Gamma(G)$ are the pairs (A, g), where $g \in G$ and A is a subset of G containing the identity element e and the element g^{-1}. The source of (A, g) is A and the range is gA. The multiplication of pairs (A, g), (B, h) in $\Gamma = \Gamma(G)$ is defined for pairs with $A = hB$, in which case we set:

$$(hB, g) \cdot (B, h) = (B, gh).$$

For a vertex x of a groupoid Γ the *isotropy group* S of x consists of all $\gamma \in \Gamma$ with $s(\gamma) = r(\gamma) = x$. Decomposing $\Gamma(G)$ in a disjoint union $\bigcup_i \Delta_i$ of connected components Δ_i we represent $K\Gamma(G)$ as a direct sum $\bigoplus_i K\Delta_i$. If Δ_i has a finite number m of vertices, then by Proposition 3.1 (b) of [2], $K\Delta_i \cong M_m(KH)$, where H is the isotropy group of some vertex of Δ_i. Then for a finite group G the groupoid algebra $K\Gamma(G)$ is isomorphic to a direct sum of algebras of form $M_m(KH)$, where H is a subgroup of G (see [2, Theorem 3.2]). Theorem 2.6 of [2], on the other hand, establishes a one-to-one correspondence between the partial representations of a finite group G and K-representations of $K\Gamma(G)$. More precisely, there is a partial representation $\lambda: G \to K\Gamma(G)$ such that every partial representation $\pi: G \to \mathrm{End}(V)$, is of form $\phi \circ \lambda$, where $\phi: K\Gamma(G) \to \mathrm{End}(V)$ is a representation of $K\Gamma(G)$. Obviously, for an arbitrary representation ϕ of $K\Gamma(G)$, $\phi \circ \lambda$ is a partial representation of G. Since $\{\lambda(g); g \in G\}$ generate the K-algebra $K\Gamma(G)$ (see [2, pp. 516–517]), this yields the desired correspondence. It can be verified that it preserves the irreducibility.

The mentioned results from [2] (except Proposition 3.1) refers to partial representations of finite groups. The main working tool is the map λ which is defined by $\lambda(g) = \sum_{A \ni g^{-1}} (A, g)$. This sum becomes infinite if we take an infinite group G. This is the main obstruction in dealing with (abstract) infinite groups and because of this fact we are not able to prove the isomorphism $K\Gamma(G) \cong K_{\mathrm{par}}(G)$. However, we replace λ by a family of maps λ_Δ, each corresponding to a connected component Δ of $\Gamma(G)$ with finite number of vertices, such that it permits to relate the finite degree partial representations of an infinite group G with representations of $K\Gamma(G)$.

For a subgroupoid Δ of $\Gamma(G)$ denote by V_Δ the set of vertices of Δ.

THEOREM 1. *Let G be a group. For every connected component Δ of $\Gamma(G)$ with finite number of vertices the map $\lambda_\Delta: G \to K\Delta$, defined by*

$$\lambda_\Delta(g) = \sum_{\substack{A \in V_\Delta \\ A \ni g^{-1}}} (A, g),$$

is a partial representation of G into $K\Delta$. Moreover, for each irreducible finite degree K-representation $\varphi: K\Delta \to \mathrm{End}(V)$, $\varphi \circ \lambda_\Delta$ is an irreducible partial representation of G. Conversely, for every irreducible finite degree partial

K-representation $\pi: G \to \mathrm{End}(V)$ there exist a unique connected component Δ of $\Gamma(G)$ with finite number of vertices and a unique irreducible representation $\tilde{\pi}: K\Delta \to \mathrm{End}(V)$ such that $\tilde{\pi} \circ \lambda_\Delta = \pi$.

Let H be a subgroup of G and V be a free right KH-module of finite rank. Observe that for free KH-modules the "finite rank" is a well defined number. If $\varphi: G \to \mathrm{End}(V_{KH})$ is a partial KH-representation of a finite group G then V becomes a $K_{\mathrm{par}}(G)$-KH-bimodule and each $K_{\mathrm{par}}(G)$-KH-bimodule gives rise to a partial KH-representation of G on V_{KH} in such a way that equivalent partial representations correspond to isomorphic $K_{\mathrm{par}}(G)$-KH-bimodules. For arbitrary G we say that $\varphi: G \to \mathrm{End}(V_{KH})$ is *monomial over H* if there is a free KH-basis for V such that for every $g \in G$ each row and each column of the matrix of $\varphi(g)$ contains at most one non-zero entry, which is an element of H (observe that we allow zero rows and zero columns).

It follows from Theorem 1 that each irreducible finite dimensional partial G-space V can be considered as an irreducible left $K\Delta$-module where Δ is a connected component of $\Gamma(G)$ with finite number of vertices.

THEOREM 2. *Let $\pi: G \to \mathrm{End}(V)$ be an irreducible finite degree partial representation of G, Δ be the connected component of $\Gamma(G)$ related to π and ${}_{K\Delta}V$ be the left $K\Delta$-module corresponding to π. Then ${}_{K\Delta}V \cong {}_{K\Delta}W \otimes_{KH} U$, where $H \subseteq G$ is the isotropy group of a vertex Δ, U is an irreducible left KH-module and W is the $K\Delta$-KH-bimodule corresponding to a monomial over H partial representation of G.*

If G is finite then, of course, each connected component of $\Gamma(G)$ has a finite number of vertices and $K\Gamma(G) \cong K_{\mathrm{par}}(G)$. Thus we have:

COROLLARY 3. *Let G be a finite group. Then every irreducible left $K_{\mathrm{par}}(G)$-module is isomorphic to ${}_{K_{\mathrm{par}}(G)}V \otimes_{KH} W$, where W is some left KH-module and ${}_{K_{\mathrm{par}}(G)}V_{KH}$ is the bimodule corresponding to a monomial over H partial representation.*

It follows from Theorem 2 that in order to reduce the description of irreducible finite degree partial representations of G to the description of (usual) irreducible representations of subgroups of G, one needs to know the monomial over subgroups partial representations determined by the connected components of $\Gamma(G)$, which have finite number of vertices.

Let H be a subgroup of an abelian group G. We give the description of monomial over H partial representations of degree ≤ 3, which come from the connected components (the authors obtained also the description of 4×4 such partial representations which are not included here because of space restriction). We denote by $e_{i,j}(h)$ the elementary matrix whose unique non-zero entry is $h \in H$, which is placed in the intersection of the i-th row and j-th column.

1×1-PARTIAL MONOMIAL REPRESENTATIONS.

$$\varphi_H\colon h \longmapsto h, \quad g \longmapsto 0, \quad \text{for each } h \in H, g \in G \setminus H.$$

2×2-PARTIAL MONOMIAL REPRESENTATIONS.

$$\begin{aligned}\varphi_{H,a}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h), \quad a^{-1}h \longmapsto e_{2,1}(h), \quad \text{for every } h \in H,\\ & g \longmapsto 0, \text{ if } g \notin H \cup aH \cup a^{-1}H.\end{aligned}$$

3×3-PARTIAL MONOMIAL REPRESENTATIONS.

$$\begin{aligned}\varphi_{H,a,b,1}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h), \quad bh \longmapsto e_{1,3}(h), \quad a^{-1}bh \longmapsto e_{2,3}(h),\\ & a^{-1}h \longmapsto e_{2,1}(h), \quad b^{-1}h \longmapsto e_{3,1}(h), \quad ab^{-1}h \longmapsto e_{3,2}(h),\\ & g \longmapsto 0, \quad \text{if } g \notin H \cup aH \cup bH \cup a^{-1}bH \cup a^{-1}H \cup b^{-1}H \cup ab^{-1}H,\\ & \text{where } a^2, b^2, (a^{-1}b)^2, ab \notin H, a^2 \notin bH, b^2 \notin aH;\\ \varphi_{H,a,b,2}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h) + e_{2,1}(h_1h), \quad bh \longmapsto e_{1,3}(h),\\ & b^{-1}h \longmapsto e_{3,1}(h), \quad abh \longmapsto e_{2,3}(h_1h), \quad ab^{-1}h \longmapsto e_{3,2}(h),\\ & g \longmapsto 0, \quad \text{if } g \notin H \cup aH \cup bH \cup abH \cup b^{-1}H \cup ab^{-1}H,\\ & \text{where } a^2 = h_1 \in H, \text{ but } b^2 \notin H \cup aH, ab \notin H;\\ \varphi_{H,a,b,3}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h) + e_{2,1}(h_1h), \quad bh \longmapsto e_{1,3}(h) + e_{3,1}(h_2h),\\ & abh \longmapsto e_{2,3}(h_1h) + e_{3,2}(h_2h),\\ & g \longmapsto 0, \quad \text{if } g \notin H \cup aH \cup bH \cup abH,\\ & \text{where } a^2 = h_1 \in H, b^2 = h_2 \in H, ab \notin H;\\ \varphi_{H,a,4}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h) + e_{3,1}(h), \quad a^{-1}h \longmapsto e_{2,1}(h) + e_{1,3}(h),\\ & a^2h \longmapsto e_{3,2}(h), \quad a^{-2}h \longmapsto e_{2,3}(h),\\ & g \longmapsto 0, \quad \text{if } g \notin H \cup aH \cup a^{-1}H \cup a^2H \cup a^{-2}H,\\ & \text{where } a^3, a^4 \notin H;\\ \varphi_{H,a,5}\colon\ & h \longmapsto hE, \quad ah \longmapsto e_{1,2}(h) + e_{3,1}(h), \quad a^{-1}h \longmapsto e_{2,1}(h) + e_{1,3}(h),\\ & a^2h \longmapsto e_{2,3}(h_1h) + e_{3,2}(h),\\ & g \longmapsto 0, \quad \text{if } g \notin H \cup aH \cup a^{-1}H \cup a^2H,\\ & \text{where } a^4 = h_1 \in H.\end{aligned}$$

THEOREM 4. *Suppose that H is a subgroup of an abelian group G. Then each irreducible monomial over H partial representation of G of degree ≤ 3, which comes from a connected component of $\Gamma(G)$, is equivalent to one of the following.*

(i) φ_H (degree 1).

(ii) $\varphi_{H,a}$ for some $a \in G \setminus H$ with $a^{-1}H \neq aH$ (degree 2).

(iii) $\varphi_{H,a,b,1}$, $\varphi_{H,a,b,2}$, $\varphi_{H,a,b,3}$, $\varphi_{H,a,4}$, $\varphi_{H,a,5}$ for some $a, b \in G \setminus H$ with $aH \neq bH$ (degree 3).

REFERENCES

1. J. Cuntz and W. Krieger, *A Class of C^*-Algebras and Topological Markov Chains.* Invent. Math. **56**(1980), 251–268.
2. M. Dokuchaev, R. Exel and P. Piccione, *Partial representations and partial group algebras.* J. Algebra (1) **226**(2000), 505–532.
3. R. Exel, *Partial Actions of Groups and Actions of Semigroups.* Proc. Amer. Math. Soc. (12) **126**(1998), 3481–3494.
4. ———, *Amenability for Fell Bundles.* J. Reine Angew. Math. **492**(1997), 41–73.
5. R. Exel and M. Laca, *Cuntz-Krieger Algebras for Infinite Matrices.* J. Reine Angew. Math. **512**(1999), 119–172.
6. J. C. Quigg and I. Raeburn, *Characterizations of Crossed Products by Partial Actions.* J. Operator Theory **37**(1997), 311–340.

Departamento de Matemática
Universidade de São Paulo
Brazil
email: dokucha@ime.usp.br

Department of Mathematics
Czech Technical University in Prague
Czech Republic
email: natalia@math.feld.cvut.cz

C. R. Math. Rep. Acad. Sci. Canada Vol. **24** (2), 2002 pp. 91–96

ON POLYNOMIALS HAVING ZEROS ON THE UNIT CIRCLE

PIROSKA LAKATOS

Presented by M. Ram Murty, FRSC

RÉSUMÉ. Nous prouvons que toute zéros du polynomial réel réciproque

$$h_m(z) = l(z^m + z^{m-1} + \cdots + z + 1) + \sum_{k=1}^{[\frac{m}{2}]} a_k(z^{m-k} + z^k) \quad (z \in \mathbb{C})$$

de degré m où $l, a_0, a_1, \ldots, a_{[\frac{m}{2}]} \in \mathbb{R}$, $l \neq 0$, $m \in \mathbb{N}$, $m \geq 2$ sont sur le cercle d'unité si $|l| \geq 2\sum_{k=1}^{[\frac{m}{2}]} |a_k|$. Utilisant ce résultat nous recevrons qu'un polynomial réciproque $P_m(z) = \sum_{j=0}^{m} A_j z^j$ $(z \in \mathbb{C})$ de degré $m \geq 2$ avec coefficients réels $A_j \in \mathbb{R}$ a zéros sur le cercle d'unité supposant que

$$|A_m| \geq \sum_{k=1}^{m-1} |A_k - A_m|.$$

1. Introduction. The spectral properties of Coxeter transformations play an important role not only in the representation theory of finite dimensional algebras (see [2]) but in number theory too [3], [4]. The investigation of spectral properties of Coxeter transformation of certain oriented graphs lead us to the polynomials

$$h_m(z) = l(z^m + z^{m-1} + \cdots + z + 1) + \sum_{k=1}^{[\frac{m}{2}]} a_k(z^{m-k} + z^k) \quad (z \in \mathbb{C})$$

where $l, a_0, a_1, \ldots, a_{[\frac{m}{2}]} \in \mathbb{R}$, $l \neq 0$, $m \in \mathbb{N}$, $m \geq 2$. Clearly the zeros of the first term of h_m are on the unit circle. We shall show that adding the linear combination of symmetric terms $\sum_{k=1}^{[\frac{m}{2}]} a_k(z^{m-k} + z^k)$ to the polynomial $l\frac{z^{m+1}-1}{z-1}$ the zeros remain on the unit circle provided that l is large enough.

Our basic tool is the Chebyshev transformation of semi-reciprocal polynomials. We recall the properties of the Chebyshev transformation (see [1]).

Received by the editors March 7, 2002.

Research partially supported by grant NSERC of Canada No. A-7257

AMS subject classification: Primary: 30C15; secondary: 12D10, 42C05.

Key words and phrases: reciprocal, semi-reciprocal polynomials, Chebyshev transform, zeros on the unit circle.

DEFINITION 1. A polynomial p of the form

$$p(z) = \sum_{j=0}^{2n} a_j z^j \quad (z \in \mathbb{C})$$

where $n \in \mathbb{N}$, $a_0, \ldots, a_{2n} \in \mathbb{R}$ and $a_j = a_{2n-j}$ $(j = 0, \ldots, n-1)$ is called a *real semi-reciprocal polynomial* of degree at most $2n$. If $a_{2n} \neq 0$ we call p a *real reciprocal polynomial* of degree $2n$.

Denote by $\mathcal{R}_{2n}$ the set of all real semi-reciprocal polynomials of degree at most $2n$.

If $p \in \mathcal{R}_{2n}$, and p is not the zero polynomial then there is an integer k, $0 \leq k \leq n$ such that $a_{2n} = a_{2n-1} = \cdots = a_{n+k+1} = 0 = a_{n-k-1} = \cdots = a_0$ but $a_{n+k} = a_{n-k} \neq 0$.

PROPOSITION 1 (*cf.* [1, PROPOSITION 2.10]). *Every non-zero polynomial $p \in \mathcal{R}_{2n}$ has the decomposition*

$$p(z) = a_{n+k} z^{n-k} \prod_{j=1}^{k} (z^2 - \alpha_j z + 1) \tag{1}$$

where $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$, $a_{n+k} \neq 0$ for some k with $0 \leq k \leq n$ and the convention $\prod_{j=1}^{0} b_j := 1$ is adopted. If $p \in \mathcal{R}_{2n}$ is a reciprocal polynomial of degree $2n$ then (1) holds with $k = n$.

DEFINITION 2. The Chebyshev transform of a non-zero polynomial $p \in \mathcal{R}_{2n}$ having the decomposition (1) is defined by

$$\mathcal{T}p(x) = a_{n+k} \prod_{j=1}^{k} (x - \alpha_j) \tag{2}$$

while for the zero polynomial p let $\mathcal{T}p(x) = 0$.

It is clear that $\mathcal{T}$ maps $\mathcal{R}_{2n}$ into the set $\mathcal{P}_n$ of all polynomials of degree $\leq n$ with real coefficients.

PROPOSITION 2 (*cf.* [1, PROPOSITION 2.10]). *The Chebyshev transform $\mathcal{T}$ is an isomorphism of the (real) vector space $\mathcal{R}_{2n}$ onto $\mathcal{P}_n$.*

The zeros of polynomials $p \in \mathcal{R}_{2n}$ and that of $\mathcal{T}p$ are connected by:

LEMMA 1. *Let p be a real reciprocal polynomial of degree $2n$. Then all zeros of p are on the unit circle if and only if all zeros of its Chebyshev transform $\mathcal{T}p$ are in the closed interval $[-2, 2]$.*

2. Results and proofs.

THEOREM 1. *All zeros of the (real reciprocal) polynomial*

$$h_m(z) = l(z^m + z^{m-1} + \cdots + z + 1) + \sum_{k=1}^{[\frac{m}{2}]} a_k(z^{m-k} + z^k) \quad (z \in \mathbb{C}) \tag{3}$$

of degree m where $l, a_0, a_1, \ldots, a_{[\frac{m}{2}]} \in \mathbb{R}$, $l \neq 0$, $m \in \mathbb{N}$, $m \geq 2$ are on the unit circle if

$$|l| \geq 2\sum_{k=1}^{[\frac{m}{2}]} |a_k|. \tag{4}$$

PROOF. If $m = 2n$ is an even number then with the notation $v_j(z) = z^j + z^{j-1} + \cdots + 1 = \frac{z^{j+1}-1}{z-1}$, $e_j(z) = z^j$, $w_j(z) = z^j + 1$ $(j = 0, 1, \ldots)$ we have

$$h_{2n}(z) = lv_{2n}(z) + \sum_{k=1}^{n} a_k e_k(z) \cdot w_{2n-2k}(z),$$

therefore

$$\mathcal{T}h_{2n}(x) = l\mathcal{T}v_{2n}(x) + \sum_{k=1}^{n} a_k \mathcal{T}(e_k \cdot w_{2n-2k})(x).$$

Factorizing v_{2n}, w_{2n-2k} we can find the Chebychev transforms $\mathcal{T}v_{2n}, \mathcal{T}(e_k \cdot w_{2n-2k})$ directly. Moreover they can be expressed by the help of the Chebychev polynomials T_n, U_n of the first and second kind (defined by $T_n(\cos x) = \cos nx$, $U_n(\cos x) = \frac{\sin(n+1)x}{\sin x}$ $(n = 0, 1, \ldots)$ see *e.g.* [5]) as

$$\mathcal{T}v_{2n}(x) = U_n\left(\frac{x}{2}\right) + U_{n-1}\left(\frac{x}{2}\right),$$
$$\mathcal{T}(e_k \cdot w_{2n-2k})(x) = 2T_{n-k}\left(\frac{x}{2}\right).$$

Thus we have

$$\mathcal{T}h_{2n}(x_j) = l\left(U_n\left(\frac{x_j}{2}\right) + U_{n-1}\left(\frac{x_j}{2}\right)\right) + \sum_{k=1}^{n} 2a_k T_{n-k}\left(\frac{x_j}{2}\right). \tag{5}$$

Let

$$x_j = 2\cos y_j \quad \text{with} \quad y_j = \frac{j+\frac{1}{2}}{2n+1}2\pi \quad (j = 0, \ldots, n)$$

then by (5)

$$\begin{aligned} \mathcal{T}h_{2n}(x_j) &= 2\left[\frac{\frac{l}{2}\sin\frac{2n+1}{2}y_j}{\sin\frac{1}{2}y_j} + \sum_{k=1}^{n} a_k \cos(n-k)y_j\right] \\ &= 2\left[\frac{\frac{l}{2}(-1)^j}{\sin\frac{y_j}{2}} + \sum_{k=1}^{n} a_k \cos(n-k)y_j\right]. \end{aligned}$$

If $j = 0, 1, \dots, n-1$ then $0 < \sin\frac{y_j}{2} < 1$, $\sum_{k=1}^{n} |a_k \cos(n-k)y_j| \le \sum_{k=1}^{n} |a_k|$ and by (4) the sign of the expression in the bracket is $(-1)^j \operatorname{sgn} l$.

If $j = n$ then $y_n = \pi$ and the expression in the bracket is

$$\frac{l}{2}(-1)^n + \sum_{k=1}^{n} a_k(-1)^{n-k} = (-1)^n \left(\frac{l}{2} + \sum_{k=1}^{n} a_k(-1)^k\right).$$

Its sign is $(-1)^n \operatorname{sgn} l$ if in (4) strict inequality holds or if in (4) we have equality and at least for one k, $(1 \le k \le n)$ we have $\operatorname{sgn} l = \operatorname{sgn}(-1)^k \operatorname{sgn} a_k$. If we have equality in (4) and $\operatorname{sgn} l = \operatorname{sgn}(-1)^{k+1} \operatorname{sgn} a_k$ for all $k = 1, \dots, n$ then the expression in the bracket is zero.

Thus either $\operatorname{sgn} \mathcal{T}h_{2n}(x_j) = \operatorname{sgn}(-1)^j \operatorname{sgn} l$ $(j = 0, \dots, n)$ or $\operatorname{sgn} \mathcal{T}h_{2n}(x_j) = \operatorname{sgn}(-1)^j \operatorname{sgn} l$ $(j = 0, 1, \dots, n-1)$ and $\mathcal{T}h_{2n}(x_n) = 0$. In both cases $\mathcal{T}h_{2n}$ has n distinct zeros in the interval $[-2, 2]$. Applying Lemma 1 we get our statement.

If $m = 2n+1$ is an odd number then we have $h_{2n+1}(z) = (z+1)\bar{h}_{2n}(z)$ with $\bar{h}_{2n}(z) = l\bar{v}_{2n}(z) + \sum_{k=1}^{n} a_k z^k \bar{w}_{2n-2k}(z)$ where

$$\bar{v}_{2n}(z) = z^{2n} + z^{2n-2} + \dots + z^2 + 1 = v_n(z^2)$$
$$\bar{w}_{2n-2k}(z) = \frac{w_{2n+1-2k}(z)}{z+1} = \frac{z^{2n+1-2k}+1}{z+1}.$$

Finding $\mathcal{T}\bar{v}_{2n}$, $\mathcal{T}(e_k \cdot \bar{w}_{2n-2k})$ from the corresponding factorizations and rewriting them by the help of Chebychev polynomials we get

$$\mathcal{T}\bar{v}_{2n}(x) = U_n\left(\frac{x}{2}\right),$$
$$\mathcal{T}(e_k \cdot \bar{w}_{2n-2k})(x) = U_{n-k}\left(\frac{x}{2}\right) - U_{n-k-1}\left(\frac{x}{2}\right).$$

where we adopt the convention $U_{-1}(x) = 0$ $(x \in \mathbb{C})$.

By the linearity of the Chebyshev transform and we have

$$\mathcal{T}\bar{h}_{2n}(x) = lU_n\left(\frac{x}{2}\right) + \sum_{k=1}^{n} a_k \left[U_{n-k}\left(\frac{x}{2}f\right) - U_{n-k-1}\left(\frac{x}{2}\right)\right]. \tag{6}$$

Let

$$\bar{x}_j = 2\cos\bar{y}_j \quad \text{with} \quad \bar{y}_j = \frac{j+\frac{1}{2}}{2n+2}2\pi \quad (j = 0, \dots, n)$$

then by (6) we have

$$\begin{aligned}\mathcal{T}\bar{h}_{2n}(\bar{x}_j) &= 2\left[\frac{l}{2}\frac{\sin(n+1)\bar{y}_j}{\sin\bar{y}_j} + \frac{\sum_{k=1}^{n} a_k \cos\frac{2n-2k+1}{2}\bar{y}_j}{2\cos\frac{\bar{y}_j}{2}}\right] \\ &= 2\frac{\frac{l}{2}(-1)^j + \sum_{k=1}^{n} a_k \sin\frac{\bar{y}_j}{2}\cos\frac{2n-2k+1}{2}\bar{y}_j}{\sin\bar{y}_j}.\end{aligned}$$

Since $\bar{y}_j \in]0, \pi[$ we have $\sin \bar{y}_j > 0$, $0 < \sin \frac{\bar{y}_j}{2} < 1$, $|\cos \frac{2n-2k+1}{2}\bar{y}_j| \leq 1$ for all $k = 1, \dots, n$ therefore the sign of the expression in the bracket is $\operatorname{sgn} l \operatorname{sgn}(-1)^j$. Thus $\operatorname{sgn}(\mathcal{T}\bar{h}_{2n}(x_j)) = \operatorname{sgn} l \operatorname{sgn}(-1)^j$ $(j = 0, 1, \dots, n)$ proving that $\mathcal{T}\bar{h}_{2n}$ has n different zeros in $[-2, 2]$. Applying Lemma 1 completes the proof. ∎

We remark that the condition $|l| \geq \sum_{k=1}^{[\frac{m}{2}]} |a_k|$ is not sufficient for all polynomials h_m to have all zeros on the unit circle. A counterexample is the polynomial $g(x) = (x^4 + x^3 + x^2 + x + 1) + 2x^2$.

Theorem 1 can be formulated in a more symmetric way. This formulation explains, in a certain way, the appearance of the factor 2 in (4).

THEOREM 2. *All zeros of the reciprocal polynomial*

$$P_m(z) = \sum_{j=0}^{m} A_j z^j \quad (z \in \mathbb{C})$$

of degree $m \geq 2$ with real coefficients $A_j \in \mathbb{R}$ (i.e., $A_m \neq 0$ and $A_j = A_{m-j}$ for all $j = 0, \dots, [\frac{m}{2}]$) are on the unit circle, provided that

$$|A_m| \geq \sum_{k=1}^{m-1} |A_k - A_m|. \tag{7}$$

PROOF. Comparing the coefficients of z^j in h_m and P_m we see that for even $m = 2n$

$$A_{2n} = A_0 = l, \ \ A_{2n-1} = A_1 = l + a_1, \dots, A_{n+1} = A_{n-1} = l + a_{n-1}, \ \ A_n = l + 2a_n$$

thus $l = A_{2n}$, $a_k = A_{2n-k} - A_{2n} = A_k - A_{2n}$ for $k = 1, 2, \dots, n-1$ and $2a_n = A_n - A_{2n}$. Therefore the condition (4) can be written as

$$|A_{2n}| \geq 2 \sum_{k=1}^{n-1} |A_k - A_{2n}| + |A_n - A_{2n}| = \sum_{k=1}^{2n-1} |A_k - A_{2n}|$$

which is the same as (7).

For odd $m = 2n + 1$ the comparison of the coefficients gives that

$$A_{2n+1} = A_0 = l, \quad A_{2n} = A_1 = l + a_1, \dots, A_{n+1} = A_n = l + a_n$$

thus $l = A_{2n+1}$, $a_k = A_{2n+1-k} - A_{2n+1} = A_k - A_{2n+1}$ for $k = 1, 2, \dots, n$ and (4) can be written as

$$\begin{aligned} |A_{2n+1}| &\geq 2 \sum_{k=1}^{n} |A_k - A_{2n+1}| = \sum_{k=1}^{n} (|A_k - A_{2n+1}| + |A_{2n+1-k} - A_{2n+1}|) \\ &= \sum_{k=1}^{2n} |A_k - A_{2n+1}| \end{aligned}$$

proving (7). ∎

REMARK 1. The manuscript of a full version of this paper was sent to Professor Andrzej Schinzel for his comments. After seeing my paper, he has informed me that he can generalize Theorem 2 to the case of self-inversive polynomials over $\mathbb{C}$ and that he intends to publish this generalization independently.

REFERENCES

1. J. W. Cannon and P. Wagreich, *Growth function of surface groups.* Math. Ann. **293**(1992), 239–257.
2. V. Dlab and C. M. Ringel, *Indecomposable representations of graphs and algebras.* Mem. Amer. Math. Soc. **173**(1976), 1–57.
3. J. F. McKee, P. Rowlinson and C. J. Smyth, *Salem numbers and Pisot numbers from stars.* In: Number theory in progress, Vol. 1 (eds. K. Győry, *et al.*), de Gruyter, Berlin, 1999, 309–319.
4. P. Lakatos, *Salem numbers, PV numbers and spectral radii of Coxeter transformations.* C. R. Math. Rep. Acad. Sci. Can. (3) **23**(2001), 71–77.
5. T. J. Rivlin, *Chebyshev polynomials.* John Wiley & Sons, New York, 1990.

Institute of Mathematics and Informatics
Debrecen University
4010 Debrecen, pf.12
Hungary
email: lapi@math.klte.hu