BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 1, Number 6, November 1979

IS COMPUTING WITH THE FINITE FOURIER TRANSFORM PURE OR APPLIED MATHEMATICS?

BY L. AUSLANDER AND R. TOLIMIERI

HISTORICAL INTRODUCTION by L. Auslander

Let me begin with my view of a bit of history.

Before the Second World War mathematics in the United States was a servant of the needs of others and mathematicians taught service courses. Indeed, while A. Weil was teaching at an Eastern university it would be only a slight exaggeration to say that he was forbidden from presenting proofs in class and was called on the carpet by a dean for breaking this structure. In the years after the War, mathematics became a subject in its own right. Proofs became acceptable, as the creation of the "new math" proved to the world. Mathematicians were in demand, were men in their own right and no one's servants.

However, this growth period had a very unfortunate side affect. While mathematics was becoming a subject in its own right, many of its practitioners wanted to rid themselves of their former servant image. They had felt denigrated by the service role; so they denigrated service mathematics. Unfortunately, they lumped together service mathematics and applied mathematics. And so during this growth period of mathematics, there sprang up a distinction between pure and applied mathematics. During these years, the applied mathematicians felt the pure mathematicians looked down on them, and so the communications between the pure and applied mathematicians virtually dried up.

In this paper we willl show that there is really not much difference between pure and applied mathematics. Indeed, we will cite instances of pure and applied mathematicians doing the same or analogous mathematics, but because of the lack of communication neither knew of the others' work.

With these broad generalities stated, let me try to explain how I came to the writing of this paper. This may perhaps serve as an example of how the gap between pure and applied mathematicians can be bridged.

I became interested in the study of the finite Fourier transform because I needed to know the eigenvalues of the finite Fourier transform. This arose in the study of the multiplicity of the regular representation of a solvmanifold. This problem was solved and the solution can be found in [**8**, p. 95]. Tolimieri, and Tolimieri and I, took up this problem in [**18**] and [**3**] and related the eigenvalue problem of the finite Fourier transform to a certain algebra of theta functions as discussed in Chapter I of this paper. I felt that

Received by the editors March 1, 1979.
AMS (MOS) subject classifications (1970). Primary 42A68; Secondary 68A20, 68A10, 10G05, 22E25.

0002-9904/79/0000-0501/$13.75

the mathematicians at the IBM Watson Research Center at Yorktown Heights, New York, might be interested in these results. They were, and they invited me to give a talk on my work. After the talk, James Cooley was kind enough to point out that two electrical engineers, McClellan and Parks [16], and the applied mathematician I. J. Good [11] had written interesting papers on this subject. I. J. Good pointed out that Gauss had studied and really solved the problem of the eigenvalues of the finite Fourier transform. All these ideas are presented in Chapter I.

My interest in the computational aspects of the finite Fourier transform was aroused by the papers J. Cooley gave me. Tolimieri and I in [3] had presented a proof of the Plancherel theorem for the reals that put the Weil-Brezin (see [19] and [7]) mapping in a central position. I felt this would yield a method for computing the finite Fourier transform. Indeed it did! It yeilded the Cooley-Tukey algorithm. This inter-relation between the Cooley-Tukey algorithm and the Weil-Brezin map is discussed in Chapter II.

All this aroused my interest in the computations of the finite Fourier transform. I spent the Fall of 1977 at the IBM Watson Research Center where I worked with S. Winograd. I have presented some of Winograd's ideas in Chapter III.

Table of Contents

Mathematical Introduction

In most applications the finite Fourier transform $F(n)$, n a positive integer, is the $n \times n$ matrix whose entry in the a row and b column, $0 \leqslant a, b < n$, is

the number

$$\frac{1}{\sqrt{n}} e^{2\pi iab/n}$$

where ab denotes the product of the 2 numbers a and b. Thus

$$F(2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$F(3) = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{2\pi i2/3} \\ 1 & e^{2\pi i2/3} & e^{2\pi i/3} \end{bmatrix},$$

$$F(n) = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{2\pi i/n} & e^{2\pi i2/n} & \cdots & e^{2\pi i(n-1)/n} \\ 1 & e^{2\pi i2/n} & e^{2\pi i4/n} & \cdots & e^{2\pi i2(n-1)/n} \\ \vdots & & & & \vdots \\ 1 & e^{2\pi i(n-1)/n} & e^{2\pi i2(n-1)/n} & \cdots & e^{2\pi i(n-1)(n-1)/n} \end{bmatrix}.$$

The main problem involving the finite Fourier transform is the following. Given a complex valued function $f(a)$, $0 \leqslant a < n$, we want to compute the function $Y(b)$, $0 \leqslant b < n$, given by

$$\begin{bmatrix} Y(0) \\ \vdots \\ Y(n-1) \end{bmatrix} = F(n) \begin{bmatrix} f(0) \\ \vdots \\ f(n-1) \end{bmatrix}.$$

We call $Y(b)$ the finite Fourier transform of the function $f(a)$ and we will abbreviate this by $Y = F(n)f$ and sometimes denote $F(n)f$ by $\hat{f}$.

However, on books on Harmonic Analysis the finite Fourier transform is defined in the following, apparently, different fashion. Let $\widehat{\mathbf{Z}/n}$ denote the group of integers mod n. Let $\mathbf{C}$ denote the complex numbers and $\mathbf{C}^{\times}(n)$ denote the multiplicative group of complex numbers $e^{2\pi ik/n}$, $0 \leqslant k < n$, or what is called the group of n roots of unity. Let $\widehat{\mathbf{Z}/n}$ denote the set of homomorphisms of $\mathbf{Z}/n$ into $\mathbf{C}^{\times}(n)$. We make $\widehat{\mathbf{Z}/n}$ into a group by defining

$$(\hat{a} + \hat{b})(a) = \hat{a}(a) \cdot \hat{b}(a), \qquad a \in \mathbf{Z}/n, \hat{a}, \hat{b} \in \widehat{\mathbf{Z}/n},$$

where multiplication is in $\mathbf{C}^{\times}(n)$. The group $\widehat{\mathbf{Z}/n}$ is called the character group of $\mathbf{Z}/n$ and is well known to be isomorphic to $\mathbf{Z}/n$. We introduce the notation $\hat{b}(a) = (a, \hat{b})$, $a \in \mathbf{Z}/n$, $\hat{b} \in \widehat{\mathbf{Z}/n}$. Let f be a complex valued function on $\mathbf{Z}/n$ or $\widehat{\mathbf{Z}/n}$. We define $|f|^2 = \Sigma_{a \in G} f(a)\overline{f(a)}$ where G is $\mathbf{Z}/n$ or $\widehat{\mathbf{Z}/n}$. We define $L^2(\mathbf{Z}/n)$ or $L^2(\widehat{\mathbf{Z}/n})$ as the set of complex valued functions on $\mathbf{Z}/n$ or $\widehat{\mathbf{Z}/n}$ with the above norm. We define the finite Fourier transform

$$F(n)\colon L^2(\mathbf{Z}/n) \to L^2(\widehat{\mathbf{Z}/n})$$

by

$$(F(n)f)(\hat{b}) = \frac{1}{\sqrt{n}} \sum_{a \in \mathbf{Z}/n} f(a)\langle a, \hat{b} \rangle. \tag{1}$$

(The multiplier $1/\sqrt{n}$ is inserted to make $F(n)$ a norm preserving linear transformation or a unitary mapping or operator.)

In order to relate these two definitions of the finite Fourier transform of a function, we have to introduce some identifications or isomorphisms. First define r: $\mathbf{Z}/n \to \mathbf{C}^{\times}(n)$ by

$$r(a) = e^{2\pi ia/n}, \qquad a \in \mathbf{Z}/n.$$

We see that r is a homomorphism because

$$r(a + b) = e^{2\pi i(a+b)/n} = e^{2\pi ia/n}e^{2\pi ib/n}.$$

Noting that $e^{2\pi ia/n}$, $0 < a < n$, is not equal to 1, we have that r is an isomorphism.

Next, define s: $\mathbf{Z}/n \to \hat{\mathbf{Z}}/n$ as follows: For $b \in \mathbf{Z}/n$ define $s(b) \in \hat{\mathbf{Z}}/n$ by the formula

$$\langle a, s(b)\rangle = e^{2\pi iab/n}, \quad \text{all } a \in \mathbf{Z}/n.$$

It is straightforward to verify that s is an isomorphism. Using s to identify $\mathbf{Z}/n$ and $\hat{\mathbf{Z}}/n$ formula (1) becomes

$$(F(n)f)(b) = \frac{1}{\sqrt{n}} \sum_{0 \leqslant a < n} f(a)e^{2\pi iab/n}, \qquad 0 \leqslant b < n. \tag{2}$$

This is the same as $F(n)f$.

As above, throughout this paper we have tried to begin with the computational version of a result or problem and only then to present the more abstract or structured version of the result.

The following is a brief chapter-by-chapter survey of the contents of this paper.

In Chapter I we study the finite Fourier transform as a linear transform, rather than as the matrix product $F(n)f$. Since $F(n)$ is a unitary operator, and, as we will show, $F(n)^4 = I$, where I is the identity map, $F(n)$ is similar to a diagonal matrix whose eigenvalues are ± 1 and $\pm i$. Hence as a linear transformation, $F(n)$ is uniquely determined by the dimension of the subspaces V_α, where V_α consists of all vectors of functions in $L^2(\mathbf{Z}/n)$ such that

$$F(n)f = \alpha f, \qquad \alpha = \pm 1, \pm i.$$

The dimension of V_α is called the multiplicity of α and the problem of finding the dimension of V_α, $\alpha = \pm 1, \pm i$, is called the multiplicity problem of the finite Fourier transform $F(n)$.

In Chapter I we survey the various results, some classical and some not so classical, that enable us to solve the multiplicity problem for $F(n)$. One of these methods shows that being able to find the trace of $F(n)$ for all n is equivalent to solving the multiplicity problem. Since the trace of $F(n)$ is the quadratic Gauss sum this shows one relation of the multiplicity problem to classical mathematics. We also discuss some recent results that link the finite Fourier transform and the theory of nil-theta functions. These results center about an algebra structure that can be associated with the collection of all finite Fourier transforms $F(n)$, $n > 0$.

CHAPTER I. THE MULTIPLICITY PROBLEM

1. The Legendre symbol, quadratic reciprocity, and the trace of the finite Fourier transform. One of the simplest invariants of a linear transformation is its trace. In this section we will define the Legendre symbol (p/q) and show that if $\mathrm{Tr}(F(n))$ denotes the trace of the finite Fourier transform then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \frac{\mathrm{Tr}(F(pq))}{\mathrm{Tr}(F(p))\mathrm{Tr}(F(q))}$$

where p and q are odd primes. Of course the celebrated result of Gauss on quadratic reciprocity states that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{[(p-1)/2][(q-1)/2]}.$$

This shows one of the connections of the finite Fourier transform with classical mathematics. In order to carry out this program, we will have to introduce a representation ρ of the group $\mathbf{Z}/n^{\times}$ of units (elements with multiplicative inverse) in the ring $\mathbf{Z}/n$ on $L^2(\mathbf{Z}/n)$.

We will now start this section with a discussion of the Legendre symbol. Let p be an odd prime. Then $\mathbf{Z}/p$ is a field having p elements and $\mathbf{Z}/p^{\times}$ consists of the nonzero elements of $\mathbf{Z}/p$ and is a cyclic group of order $p-1$. Let

$$S = \{\xi^2 | \xi \in \mathbf{Z}/p^{\times}\}.$$

It is easily verified that S is a subgroup of $\mathbf{Z}/p^{\times}$. The elements of S are called quadratic residues mod p. Again one verifies that the order of the quotient group $\mathbf{Z}/p^{\times}/S$ is 2, or, for $h \in \mathbf{Z}/p^{\times}$, $h \notin S$, $\mathbf{Z}/p^{\times} = S \cup hS$ and $S \cap hS$ is empty. Let $\{1, -1\}$ be the multiplicative group of order 2 and let h be the group homomorphism of $\mathbf{Z}/p^{\times}$ onto $\{1, -1\}$ with kernel S. For the integers $\mathbf{Z}$ let $\overset{*}{p}\colon \mathbf{Z} \to \mathbf{Z}/p$ be the homomorphism with kernel consisting of the multiples of p. For $n \in \mathbf{Z}$ we define

$$\left(\frac{n}{p}\right) = \begin{cases} 0 & \text{if } n \equiv 0 \bmod p, \\ h\left(\overset{*}{p}(n)\right) & \text{if } n \not\equiv 0 \bmod p. \end{cases}$$

We call (n/p) the Legendre symbol. Since, if $n_1 \equiv n_2 \bmod p$, $(n_1/p) = (n_2/p)$ we can consider (n/p) for $n \in (\mathbf{Z}/p)^{\times}$.

Our first task is to obtain an analytic formula for (n/p).

LEMMA I.1.1. *If* $n \not\equiv 0 \bmod p$

$$\sum_{0<\xi<p} e^{2\pi i n\xi^2/p} = \left(\frac{n}{p}\right) \sum_{0<\xi<p} e^{2\pi i \xi^2/p}. \tag{1}$$

PROOF. We will call $R_p \subset \mathbf{Z}$ a complete residue system mod p if the homomorphism $\overset{*}{p}$ restricted to R_p defines a 1-1 surjection of R_p onto $\mathbf{Z}/p$. If R_p is any complete residue system mod p, then it is easy to verify that

$$\sum_{0<\xi<p} e^{2\pi i n\xi^2/p} = \sum_{\xi \in R_p} e^{2\pi i n\xi^2/p}.$$

Let $k \not\equiv 0 \bmod p$ and let R_p be a complete residue system mod p and

$$kR_p = \{k\xi | \xi \in R_p\}.$$

Then it is easily verified that kR_p is a complete residue system mod p.

Let $(n/p) = 1$ and $n \equiv k^2 \bmod p$, $k \in \mathbf{Z}/p^\times$. Then

$$\sum_{\xi \in R_p} e^{2\pi i n\xi^2/p} = \sum_{\xi \in R_p} e^{2\pi i (k\xi)^2/p} = \sum_{\eta \in kR_p} e^{2\pi i \eta^2/p} = \sum_{\xi \in R_p} e^{2\pi i \xi^2/p}$$

and we have proven this case of the lemma.

Let $(n/p) = -1$ or $p(n) \in \mathbf{Z}/p^\times$, $p(n) \notin S$. As ξ runs over a complete residue system mod p, $p(\xi^2)$ runs over S twice and the point $\{0\}$ once in $\mathbf{Z}/p$. Similarly, $h\xi^2$ runs over hS twice and the point $\{0\}$ once in $\mathbf{Z}/p$. Hence

$$\sum_{\xi \in R_p} e^{2\pi i n\xi^2/p} + \sum_{\xi \in R_p} e^{2\pi i \xi^2/p} = 2 \sum_{\xi \in R_p} e^{2\pi i \xi/p}.$$

The right-hand sum is well known to be zero and so

$$\sum_{\xi \in R_p} e^{2\pi i n\xi^2/p} = (-1) \sum_{\xi \in R_p} e^{2\pi i \xi^2/p} = \left(\frac{n}{p}\right) \sum_{\xi \in R_p} e^{2\pi i \xi^2/p}$$

and we have verified our assertion.

Let f be a function defined on complete residue systems mod n such that for $\xi \in R_n$ and $\xi' \in R_n'$ with $\xi \equiv \xi' \bmod n$ we have $f(\xi) = f(\xi')$. We will talk of f as a function on $\mathbf{Z}/n$ and use the notation $f(\xi)$, $\xi \in \mathbf{Z}/n$.

We now begin the task of introducing the representation ρ of $\mathbf{Z}/p^\times$ on $L^2(\mathbf{Z}/p)$ that combines with $F(p)$ to give another formulation of Lemma I.1.1. We begin the process of defining ρ by looking a little more closely at $L^2(\mathbf{Z}/n)$.

Let $\mathcal{F}(n)$ denote the complex valued functions on $\mathbf{Z}/n$. Let

$$f_\alpha(\beta) = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta, \end{cases} \qquad \alpha, \beta \in \mathbf{Z}/n.$$

Clearly, $\mathcal{F}(n)$ is an n-dimensional complex vector space and the n functions f_α, $\alpha \in \mathbf{Z}/n$, determine a basis of $\mathcal{F}(n)$. We will now make $\mathcal{F}(n)$ an inner product space as follows: For $f, g \in \mathcal{F}(n)$ define

$$\langle f, g \rangle = \sum_{\alpha \in \mathbf{Z}/n} f(\alpha)\bar{g}(\alpha)$$

where the bar denotes the complex conjugate. The resulting Hermitian inner product space is denoted by $L^2(\mathbf{Z}/n)$ and the n functions f_α, $\alpha \in \mathbf{Z}/n$, define an orthonormal basis of $L^2(\mathbf{Z}/n)$.

We will now describe the unitary representation ρ of $\mathbf{Z}/n^\times$ on $L^2(\mathbf{Z}/n)$. Since $\mathbf{Z}/n^\times$ acting on $\mathbf{Z}/n$ by multiplication produces a group of automorphisms of the additive group $\mathbf{Z}/n$, we may define for each $a \in \mathbf{Z}/n^\times$ a linear transformation $\rho(a)$ of $L^2(\mathbf{Z}/n)$ by setting

$$\rho(a)(f)s = f(as), \qquad s \in \mathbf{Z}/n, f \in L^2(\mathbf{Z}/n).$$

Because $\rho(a)f_\alpha = f_{a^{-1}\alpha}$, where f_α, $\alpha \in \mathbf{Z}/n$, is the orthonormal basis defined above, it follows that $\rho(a)$ is a unitary operator on $L^2(\mathbf{Z}/n)$, $a \in \mathbf{Z}/n^\times$. If $U(n)$ denotes the group of unitary operators on $L^2(\mathbf{Z}/n)$ it is easily verified

that

$$\rho\colon \mathbf{Z}/n^{\times} \to U(n)$$

is a group monomorphism.

We will now review briefly, and then extend, the material on the finite Fourier transform that we gave in the Introduction.

Let $\mathbf{C}$ denote the field of complex numbers and let $\mathbf{C}_1^{\times}$ denote the multiplicative group of complex numbers of absolute value 1. A character on $\mathbf{Z}/n$ is a group homomorphism $\lambda\colon \mathbf{Z}/n \to \mathbf{C}_1^{\times}$. The set of all characters on $\mathbf{Z}/n$ is denoted by $\widehat{\mathbf{Z}/n}$. For $\lambda_1, \lambda_2 \in \widehat{\mathbf{Z}/n}$, we define

$$(\lambda_1 + \lambda_2)a = \lambda_1(a)\lambda_2(a), \qquad a \in \mathbf{Z}/n.$$

Then $\widehat{\mathbf{Z}/n}$ is a group, isomorphic to $\mathbf{Z}/n$. For $0 \leqslant \alpha < n$, let $\lambda_\alpha\colon \mathbf{Z}/n \to \mathbf{C}_1^{\times}$ be defined by

$$\lambda_\alpha(a) = e^{-2\pi i a\alpha/n}, \qquad a \in \mathbf{Z}/n.$$

Clearly λ_α is a well defined character on $\mathbf{Z}/n$ and it is easily verified that $\widehat{\mathbf{Z}/n}$ consists of the n characters λ_α, $0 \leqslant \alpha < n$. Notice $\lambda_\alpha \in L^2(\mathbf{Z}/n)$ and

$$\langle \lambda_\alpha, \lambda_\beta \rangle = \begin{cases} n, & \alpha = \beta, \\ 0, & \alpha \neq \beta, \end{cases}$$

$\alpha, \beta \in \mathbf{Z}/n$. Hence the λ_α, $0 \leqslant \alpha < n$, are an orthogonal basis of $L^2(\mathbf{Z}/n)$. We can now define the finite Fourier transform $F(n)$ of $\mathbf{Z}/n$ as the linear mapping of $L^2(\mathbf{Z}/n)$ defined for $f \in L^2(\mathbf{Z}/n)$ by

$$(F(n)f)(\alpha) = \frac{1}{\sqrt{n}}\langle f, \lambda_\alpha \rangle = \frac{1}{\sqrt{n}} \sum_{\beta \in \mathbf{Z}/n} f(\beta) e^{2\pi i \alpha\beta/n}$$

where $\alpha \in \mathbf{Z}/n$. Notice that

$$F(n)f_\alpha = \frac{1}{\sqrt{n}}\lambda_{-\alpha} \quad \text{and} \quad F(n)\lambda_\alpha = \sqrt{n}\, f_\alpha,$$

thus

$$F(n)^2 f_\alpha = f_{-\alpha} \quad \text{and} \quad F(n)^4 = I$$

where I is the identity mapping. Also

$$\langle F(n)f_\alpha, F(n)f_\beta \rangle = \frac{1}{n}\langle \lambda_{-\alpha}, \lambda_{-\beta} \rangle = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

Hence $\langle F(n)f_\alpha, F(n)f_\beta \rangle = \langle f_\alpha, f_\beta \rangle$ and $F(n)$ is a unitary operator on $L^2(\mathbf{Z}/n)$ of order 4. We also have that

$$(F(n)^{-1}f)(\alpha) = \frac{1}{\sqrt{n}}\langle f, \lambda_{-\alpha} \rangle = \frac{1}{n} \sum_{\beta \in \mathbf{Z}/n} f(\beta) e^{-2\pi i \alpha\beta/n}.$$

Since

$$F(n)f_\beta(\alpha) = \frac{1}{\sqrt{n}}\langle f_\beta, \lambda_\alpha \rangle = \frac{1}{\sqrt{n}} \sum_{\gamma \in \mathbf{Z}/n} e^{2\pi i \gamma\beta/n} f_\gamma(\alpha)$$

it follows that the matrix of $F(n)$ with respect to the basis f_β, $\beta \in \mathbf{Z}/n$, is

given by

$$\frac{1}{\sqrt{n}}(e^{2\pi i\beta\gamma/n}), \qquad 0 \leqslant \beta,\gamma < n.$$

Although the linear transformation $F(n)$ is represented by different matrices relative to different bases we will denote the above matrix also by $F(n)$.

Since $\rho(a) \in U(n)$, $a \in \mathbf{Z}/n^{\times}$, and $F(n) \in U(n)$ we can form $F(n)\rho(a)F(n)^{-1}$. Because $\rho(a)\lambda_{\alpha} = \lambda_{a\alpha}$ we have immediately that

$$F(n)\rho(a)F(n)^{-1} = \rho(a)^{-1} = \rho(a^{-1}).$$

We can now relate the above results to Lemma I.1.1. and the results of Gauss on quadratic reciprocity and the value of quadratic Gauss sums.

If Tr( ) denotes the trace of the linear transformation in the bracket we have immediately that

$$\mathrm{Tr}(F(n)) = \sum_{\alpha\in\mathbf{Z}/n} e^{2\pi i\alpha^2/n},$$

$$\mathrm{Tr}(\rho(a)F(n)) = \sum_{\alpha\in\mathbf{Z}/n} e^{2\pi ia\alpha^2/n}.$$

Now let n and m be relatively prime positive integers and let $\gamma = \alpha m + \beta n$, $0 \leqslant \alpha < n, 0 \leqslant \beta < m$.

The Chinese remainder theorem implies that the set of all such γ is a complete residue system mod nm. Now let $f_{\alpha} \in L^2(\mathbf{Z}/n)$, $0 \leqslant \alpha < n$, and let $f_{\beta} \in L^2(\mathbf{Z}/n)$, $0 \leqslant \beta < m$, be the basis of $L^2(\mathbf{Z}/n)$ and $L^2(\mathbf{Z}/m)$ as defined above. Let $\chi(f_{\alpha}, f_{\beta}) = f_{\alpha m+\beta n}$, where f_{γ}, $0 \leqslant \gamma < nm$, is a basis of $L^2(\mathbf{Z}/nm)$ as above. Extend χ to a bilinear mapping of $L^2(\mathbf{Z}/n) \times L^2(\mathbf{Z}/m) \to L^2(\mathbf{Z}/nm)$. Then χ induces a linear map χ^*: $L^2(\mathbf{Z}/n) \otimes L^2(\mathbf{Z}/m) \to L^2(\mathbf{Z}/nm)$. It is easily verified that χ^* is an isomorphism. A straightforward computation then shows that the tensor product of the linear operators $\rho(m)F(n)$ and $\rho(n)F(m)$ satisfies

$$\rho(m)F(n) \otimes \rho(n)F(m) = F(nm). \tag{2}$$

The result of Lemma I.1.1 can now be stated as

$$\mathrm{Tr}(\rho(h)F(p)) = \left(\frac{h}{p}\right)\mathrm{Tr}(F(p)), \qquad h \not\equiv 0 \bmod p, \tag{3}$$

p an odd prime. Since the trace of a tensor product is the product of the traces of the factors, we have from equations (2) and (3) that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)\mathrm{Tr}(F(p))\mathrm{Tr}(F(q)) = \mathrm{Tr}(F(pq))$$

or

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \mathrm{Tr}\frac{\mathrm{Tr}(F(pq))}{\mathrm{Tr}(F(p))\mathrm{Tr}(F(q))} \tag{4}$$

and we have verified the result we stated at the beginning of this section.

The formula

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{[(p-1)/2][(q-1)/2]}, \qquad p \text{ and } q \text{ odd primes}, \tag{5}$$

is Gauss' celebrated formula for quadratic reciprocity. Formula (5) has many elementary proofs; see for instance, Hardy and Wright [12]. But Gauss also established the following fundamental result.

THEOREM I.1.2. *Let $F(n)$ denote the finite Fourier transform on $L^2(\mathbf{Z}/n)$ and let* $\mathrm{Tr}(F(n))$ *denote the trace of $F(n)$. Then*

$$\mathrm{Tr}(F(n)) = \begin{cases} i+1 & \text{if } n \equiv 0 \bmod 4, \\ 1 & \text{if } n \equiv 1 \bmod 4, \\ 0 & \text{if } n \equiv 2 \bmod 4, \\ i & \text{if } n \equiv 3 \bmod 4. \end{cases} \tag{6}$$

Although formula (5) has elementary proofs Theorem I.1.2 has, to our knowledge no elementary proof and seems to be much deeper than quadratic reciprocity. In [13] there is an interesting discussion of the many proofs of quadratic reciprocity.

In the next section we will relate the problem of computing $\mathrm{Tr}(F(n))$, $n > 0$, to the multiplicity problem for $F(n)$. However, before doing this, we pause to show what insights elementary considerations can give us about Theorem I.1.2.

We begin by showing that $\mathrm{Tr}(F(2r)) = 0$, r odd, is easily verified. For

$$\mathrm{Tr}(F(2r)) = \sum_{0<\xi<r} e^{2\pi i\xi^2/2r} + \sum_{0<\xi<r} e^{2\pi i(\xi+r)^2/2r}.$$

Because r is odd, $e^{2\pi ir/2} = -1$ and $e^{2\pi i(\xi+r)^2/2r} = -1e^{2\pi i\xi^2/2r}$ and we have established that $\mathrm{Tr}(F(2r)) = 0$, r odd.

Let p be an odd prime. We will now show that $\mathrm{Tr}(F(4p)) = 1 + i$ implies

$$\mathrm{Tr}(F(p)) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod 4, \\ i & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

By (2)

$$\mathrm{Tr}(\rho(4)F(p))\mathrm{Tr}(\rho(p)F(4)) = \mathrm{Tr}(F(4p)) = 1 + i.$$

Lemma I.1.1 gives

$$\mathrm{Tr}(\rho(4)F(p)) = (4/p)\mathrm{Tr}(F(p)) = \mathrm{Tr}(F(p))$$

as it is easy to verify that $(4/p) = 1$. Thus

$$\mathrm{Tr}(F(p)) = \frac{1+i}{\mathrm{Tr}(\rho(p)F(4))}.$$

But we can easily write out the four terms of the sum $\mathrm{Tr}(\rho(p)F(4))$ to prove that

$$\mathrm{Tr}(\rho(p)F(4)) = \begin{cases} 1+i & \text{if } p \equiv 1 \bmod 4, \\ 1-i & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

This shows that

$$\mathrm{Tr}(F(p)) = \begin{cases} 1 & \text{if } n \equiv 1 \bmod 4, \\ i & \text{if } n \equiv 3 \bmod 4. \end{cases}$$

2. Equivalence of the trace and eigenvalue problems for the finite Fourier transform. This section relates Theorem I.2.2 or the computation of $\mathrm{Tr}(F(n))$, $n > 0$, and the multiplicity problem as defined in the Introduction for $F(n)$. Since the trace of a linear transformation is the sum of is eigenvalues, it is clear that a solution of the multiplicity problem for $F(n)$ implies Theorem I.1.2 or Gauss' result in quadratic Gauss sums. What is very surprising is that knowing $\mathrm{Tr}(F(n))$ for all n enables us to solve the multiplicity problem for $F(n)$. The proof rests on a simple observation that is at the heart of Schur's proof of the computation of $\mathrm{Tr}(F(n))$ (see [6, p. 351]). Schur observed that

$$F^2(n) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ & & & & & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 0 & 1 & 0 & \cdot & \cdot & 0 \end{bmatrix}$$

and hence the characteristic polynomial of $F^2(n)$ is given by

$$\begin{aligned} &(t-1)^{(n+1)/2}(t+1)^{(n-1)/2}, && n \text{ odd}, \\ &(t-1)^{(n+2)/2}(t+1)^{(n-2)/2}, && n \text{ even}. \end{aligned} \tag{7}$$

Since the eigenvalues of $F^2(n)$ are the square of the eigenvalues of $F(n)$, formula (7) yields both that the possible eigenvalues of $F(n)$ are ± 1, $\pm i$ and the following result on multiplicity: If $F(n)$ has m_1 eigenvalues 1, m_2 eigenvalues -1, m_3 eigenvalues i and m_4 eigenvalues $-i$ then

$$\begin{aligned} m_1 + m_2 &= \frac{n+1}{2}, \quad m_3 + m_4 = \frac{n-1}{2}, && n \text{ odd}, \\ m_1 + m_2 &= \frac{n+2}{2}, \quad m_3 + m_4 = \frac{n-2}{2}, && n \text{ even}. \end{aligned} \tag{8}$$

and

$$\mathrm{Tr}(F(n)) = (m_1 - m_2) + i(m_3 - m_4). \tag{9}$$

Now let $\mathrm{Tr}(F(n)) = \alpha + i\beta$ then (8) and (9) combine to yield for odd n

$$\begin{aligned} m_1 - m_2 &= \alpha, & m_3 - m_4 &= \beta, \\ m_1 + m_2 &= \frac{n+1}{2}, & m_3 + m_4 &= \frac{n-1}{2} \end{aligned}$$

and a similar set of equations for even n. Clearly, we can solve for m_1, m_2, m_3, m_4 in terms of n, α, β and hence once we have evaluated $\mathrm{Tr}(F(n))$ we have a complete solution for the eigenvalue problem. This shows that the trace problem for $F(n)$ is equivalent to solving the eigenvalue problem.

Theorem I.1.2 can now be stated in the following equivalent form.

THEOREM I.1.2′. *Let $F(n)$ denote the finite Fourier transform on $\mathbf{Z}/n$ and let $m_j, j = 1, 2, 3, 4$, denote the multiplicities of the eigenvalues 1, -1, i, $-i$ of $F(n)$, respectively. The value of m_j, $j = 1, 2, 3, 4$, as a function of n is given by the following table.*

n	$m_1 = 1$	$m_2 = -1$	$m_3 = i$	$m_4 = -i$
$4m$	$m+1$	m	m	$m-i$
$4m+1$	$m+1$	m	m	m
$4m+2$	$m+1$	$m+1$	m	m
$4m+3$	$m+1$	$m+1$	$m+1$	m

(10)

3. The algebra of the finite Fourier transform. Theorem I.3.1 stated below was first discovered and proven in [4] using nilpotent harmonic analysis and its proof is independent of Theorem I.1.2′. However the equivalence of Theorems I.1.2′ and I.3.1 is easily established and requires no nilpotent harmonic analysis.

Let $\mathbf{C}[X_1, X_2, X_3]$ be the polynomial algebra in three indeterminants over the complex numbers $\mathbf{C}$ and let $\mathbf{C}[X_1, X_2^2, X_3^3]$ be the subalgebra generated by X_1, X_2^2, X_3^3. Let

$$\mathfrak{A}_1 = \mathbf{C}[X_1, X_2^2, X_3^3]/(X_3^6 + X_2^6),$$
$$\mathfrak{A}_2 = \mathbf{C}[X_1, X_2^2, X_3^3]/(X_3^6 + X_1^4 X_2^2 + X_2^6),$$
$$\mathfrak{A}_3 = \mathbf{C}[X_1, X_2^2, X_3^3]/(X_3^6 + X_1^4 X_2^2),$$

where $(\ )$ denotes the principal ideal in $\mathbf{C}[X_1, X_2^2, X_3^3]$ of the polynomial in the bracket. Let $Y_1^2\, Y_2^{2b}\, Y_3^{3c}$, $a, b, c \in \mathbf{Z}$, $a, b, c \geqslant 0$, denote the image in $\mathfrak{A}_\alpha$, $\alpha = 1, 2, 3$, of the monomial $X_1^a\, X_2^{2b}\, X_3^{3c}$. Then it is easily established that

$$Y_1^a\, Y_2^{2b}\, X_3^{3c}, \qquad c = 0, 1, a, b > 0,$$

is a vector space basis of the algebra $\mathfrak{A}_\alpha$, $\alpha = 1, 2, 3$. It is not difficult, using elementary methods, to prove that

$$X_3^6 + X_2^6, \quad X_2^6 + X_1^4 X_2^2 + X_2^6, \quad X_2^6 + X_1^4 X_2^2$$

are each irreducible in $\mathbf{C}[X_1, X_2^2, X_3^3]$. Hence each of the algebras $\mathfrak{A}_\alpha$, $\alpha = 1, 2, 3$, has no divisors of zero.

THEOREM I.3.1. *Let $\mathfrak{A}_\alpha$, $\alpha = 1, 2, 3$, be as defined above and let $\mathfrak{F}\mathfrak{A}_\alpha \to \mathfrak{A}_\alpha$, $\alpha = 1, 2, 3$, be the linear transformation of $\mathfrak{A}_\alpha$ such that*

$$\mathfrak{F}(Y_1^a\, Y_2^{2b}\, Y_3^{3c}) = (-1)^b i^c Y_1^a\, Y_2^{2b}\, Y_3^{3c}, \qquad c = 0, 1;\ a, b \geqslant 0.$$

Let $V(n)$ be the vector subspace of $\mathfrak{A}_\alpha$ spanned by $Y_1^a\, Y_2^{2b}\, Y_3^{3c}$ where $a + 2b + 3c = n$ and let $\mathfrak{F}(n) = \mathfrak{F}|V(n)$. Then $\dim V(n) = n$ *and $\mathfrak{F}(n)$ is equivalent to the finite Fourier transform $F(n)$. Further*

$$\mathfrak{F}(fg) = \mathfrak{F}(f)\mathfrak{F}(g), \qquad f, g \in \mathfrak{A}_\alpha.$$

PROOF. Let us begin by verifying the last assertion of the theorem. To do this we need only verify it for the basis elements. Now in $\mathfrak{A}_2$ (the other cases are handled similarly)

$$Y_1^a\, Y_2^{2b} = Y_1^d\, Y_2^{2e} = Y_1^{a+d}\, Y_2^{2(b+c)},$$
$$Y_1^a\, Y_2^{2b}\, Y_3^3 \cdot Y_1^d\, Y_2^{2e} = Y_1^{a+d}\, Y_2^{2(b+c)}\, Y_3^3,$$
$$Y_1^a\, Y_2^{2b}\, Y_3^3 \cdot Y_1^d\, Y_2^{2e}\, Y_3^3 = -Y_1^{a+b+4}\, Y_2^{2(b+c+1)} - Y_1^{a+b}\, Y_2^{2(b+c+3)}$$

from which it is easy to verify that

$$\mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c} \cdot Y_1^d Y_2^{2e} Y_3^{3f}) = \mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c})\mathfrak{F}(Y_1^d Y_2^{2e} Y_3^{3f})$$

and so $\mathfrak{F}(fg) = \mathfrak{F}(f)\mathfrak{F}(g)$.

We will now indicate the inductive proof used to prove that dim $V(n) = n$ and $\mathfrak{F}(n)$ and $F(n)$ have the same eigenvalues with the same multiplicities. This will prove that $\mathfrak{F}(n)$ and $F(n)$ are equivalent. We will adopt the notation $V_\chi(n)$, $\chi = \pm 1, \pm i$, for the subspace of $V(n)$ of eigenvectors with eigenvalue χ of $\mathfrak{F}(n)$.

We must now prove that dim $V_\chi(n)$ satisfy the table of Theorem I.1.2′. This is merely a property of $\mathfrak{A}_\alpha$ and $\mathfrak{F}(n)$ and does not use Theorem I.1.2′ in its proof.

By inspection the above assertion is true for $n \leqslant 4$. Consider $n = 4m$, $m \geqslant 1$. Now

$$\mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c}) = i\, Y_1^a Y_2^{2b} Y_3^{3c}$$

if and only if $c = 1$, b is even and $a + 2b = 4m - 3$. Thus dim $V_i(4m) =$ dim $V_1(4m - 3) =$ dim $V_1(4(m - 1) + 1)$ which by induction is equal to m. Thus

$$\dim V_i(4m) = m.$$

Similarly

$$\mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c}) = -i\, Y_1^a Y_2^{2b} Y_3^{3c}$$

if and only if $c = 1$, b is odd and $a + 2b = 4m - 3$. Thus

$$\dim V_{-i}(4m) = \dim V_{-1}(4m - 3) = \dim V_{-1}(4(m - 1) + 1)$$

which by induction is equal to $m - 1$. Hence

$$\dim V_{-i}(4m) = m - 1.$$

Next

$$\mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c}) = -Y_1^a Y_2^{2b} Y_3^{3c}$$

if and only if $c = 0$, b is odd and $a + 2b = 4m$. Let $b' = b - 1$. Then b' is even and $a + 2b' = 4m - 2 = 4(m - 1) + 2$. Thus

$$\dim V_{-1}(4m) = \dim V_1(4(m - 1) + 2)$$

which by induction equals m. Thus

$$\dim V_{-1}(4m) = m.$$

Finally

$$\mathfrak{F}(Y_1^a Y_2^{2b} Y_3^{3c}) = Y_1^a Y_2^{2b} Y_3^{3c}$$

if and only if $c = 0$, b is even and $a + 2b = 4m$. These constraints are satisfied only by the following values of a and b:

$$a = 4m, \quad b = 0; \qquad a = 4(m - 1), \quad b = 2, \ldots; \qquad a = 0, \quad b = 2m.$$

Hence

$$\dim V_1(4m) = m + 1.$$

Since

$$\sum_{\chi} \dim V_{\chi}(4m) = \dim V(4m)$$

we have $\dim V(4m) = 4m$ and we have proven our result for $n = 4m$. The other cases are proven in a similar way and the proof is omitted.

Thus Theorem I.1.2′ and I′.4.1 are equivalent.

4. Direct solutions of trace and eigenvalue problems. Gauss' original proof of Theorem I.1.2 can be found in H. Rademacher [**17**]. Gauss' proof is algebraic and very different from the proofs of his theorem that one finds in the usual texts on elementary number theory. We will not summarize Gauss' proof, but we do seriously suggest that all readers examine Rademacher's account of this remarkable achievement.

We know two different analytic proofs of Theorem I.1.2 and these are the proofs most often found in elementary texts. Schur's proof was the first proof that stressed the role of the finite Fourier transform and its eigenvalues. Accordingly, we will begin with a brief discussion of Schur's proof. We will follow it with McClellan and Park's proof of Theorem I.1.2′. We will then discuss the two analytic proofs of Theorem I.1.2.

We will present Schur's proof of Theorem I.1.2 only when n is an odd prime p as this makes the discussion simpler, but exhibits all the most interesting aspects of the method of proof. For those who want the whole story, this can be found in [**6**].

As in §I.2, let m_1, m_2, m_3, m_4 denote the multiplicity for $F(p)$ of the eigenvalues $1, -1, i, -i$. In §I.2 we showed that

$$m_1 + m_2 = \frac{p+1}{2}, \qquad m_3 + m_4 = \frac{p-1}{2}$$

and

$$\operatorname{Tr}(F(p)) = m_1 - m_2 + i(m_3 - m_4).$$

Schur's method of proof is simply to obtain enough relations amongst the m's to enable them to be computed.

We begin with a result that shows that it is the sign that is the difficult part of Theorem I.1.2.

LEMMA I.4.1. *Let $F(p)$ denote the finite Fourier transform on $L^2(\mathbf{Z}/p)$, p an odd prime. Then*

$$\operatorname{Tr}(F(p)) = \begin{cases} \pm 1 & \text{if } p \equiv 1 \bmod 4, \\ \pm i & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

PROOF. A multiplicative character λ of $\mathbf{Z}/p^{\times}$ is a homomorphism of the multiplicative group $\mathbf{Z}/p^{\times}$ into $\mathbf{C}_1^{\times}$. We extend λ to a function on $\mathbf{Z}/p$ by defining $\lambda(0) = 0$. Now let λ be a nontrivial multiplicative character of $\mathbf{Z}/p^{\times}$; i.e., there exists $\xi \in \mathbf{Z}/p^{\times}$ such that $\lambda(\xi) \neq 1$. View λ as an element of $L^2(\mathbf{Z}/p)$. Since the characters

$$\lambda_{\alpha}(\xi) = e^{2\pi i \alpha \xi / p}, \qquad \alpha \in \mathbf{Z}/p,$$

determine an orthogonal basis of $L^2(\mathbf{Z}/p)$ and $\langle \lambda_{\alpha}, \lambda_{\alpha} \rangle = p$, we can write

$\lambda = \Sigma_{\alpha \in \mathbf{Z}/p} a_\alpha \lambda_\alpha$, where $a_\alpha = \langle \lambda, \lambda_\alpha \rangle / p$ or

$$a_\alpha = \frac{1}{p} \sum_{\xi \in \mathbf{Z}/p} \lambda(\xi) e^{2\pi i \alpha \xi / p}.$$

An argument similar to that given in Theorem I.1.1 gives

$$\mathrm{Tr}(F(p)) = \frac{1}{\sqrt{p}} \sum_{\xi \in \mathbf{Z}/p} \left(\frac{\xi}{p}\right) e^{2\pi i \xi / p}.$$

Thus, $\mathrm{Tr}(F(p)) = \sqrt{p}\, a_1$ where

$$a_1 = \frac{1}{p} \left\langle \left(\frac{-}{p}\right), \lambda_1 \right\rangle$$

and $(\ /p)$ is the Legendre symbol that was defined as a multiplicative character.

Again let λ be an arbitrary multiplicative character. For $c \not\equiv 0 \bmod p$,

$$\lambda(cx) = \sum_{\alpha \in \mathbf{Z}/p} a_\alpha \lambda_\alpha(cx).$$

Since $\lambda(cx) = \lambda(c)\lambda(x)$ we have

$$\sum_{\alpha \in \mathbf{Z}/p} a_\alpha \lambda_\alpha(cx) = \sum_{\alpha \in \mathbf{Z}/p} \lambda(c) a_\alpha \lambda_\alpha(x).$$

Now $\lambda_\alpha(cx) = \lambda_{c\alpha}(x)$ and so if $\alpha = c^{-1}\beta$ we have

$$\begin{aligned} \sum_{\alpha \in \mathbf{Z}/p} a_\alpha \lambda_\alpha(cx) &= \sum_{\alpha \in \mathbf{Z}/p} a_\alpha \lambda_{c\alpha}(x) \\ &= \sum_{\beta \in \mathbf{Z}/p} a_{c^{-1}\beta} \lambda_\beta(x) = \sum_{\beta \in \mathbf{Z}/p} \lambda(c) a_\beta \lambda_\beta(x). \end{aligned}$$

Hence $a_{c-\beta} = \lambda(c) a_\beta$ for all β. Hence $|a_1| = |a_c|$, where $|\ |$ denotes the absolute value, $c \not\equiv 0 \bmod p$. Thus

$$\langle \lambda, \lambda \rangle = p(p-1)|a_1|^2 = p - 1$$

and

$$|a_1| = 1/\sqrt{p}\,.$$

Since $\mathrm{Tr}(F(p)) = \sqrt{p}\, a_1$ when $\lambda = (\bar{p})$, we have

$$|\mathrm{Tr}(F(p))| = 1.$$

It is an elementary fact that $(-1/p) = (-1)^{(p-1)/2}$. By Theorem I.1.1

$$\mathrm{Tr}(F(p))^- = (-1/p)\mathrm{Tr}(F(p))$$

where the bar denotes the complex conjugate. Hence

$$\mathrm{Tr}(F(p))^2(-1/p) = 1$$

or

$$\mathrm{Tr}(F(p)) = \begin{cases} \pm 1 & \text{if } p \equiv 1 \bmod 4, \\ \pm i & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

Lemma I.4.1 implies that

$$m_1 - m_2 = \pm 1, \quad m_3 = m_4 \qquad \text{if } p \equiv 1 \bmod 4,$$
$$m_1 = m_2, \qquad m_3 - m_4 = \pm 1 \quad \text{if } p \equiv 3 \bmod 4.$$

The fact that $\det(F(p)) = (-1)^{m_2} i^{m_3} (-i)^{m_4}$ will enable us to obtain the last relation we require.

Let

$$A = \det(e^{2\pi i \alpha \xi / p})_{0 < \alpha, \xi < p}$$

then

$$A = p^{p/2} \det(F(p))$$

and A is a Vandermond determinant. Hence

$$A = \prod_{0 \leq s, r < p} (e^{2\pi i r/p} - e^{2\pi i s/p}).$$

Let $\eta = e^{\pi i/p}$; we have

$$A = \prod_{0 \leq s, r < p} \eta^{r+s} (\eta^{r-s} - \eta^{-(r-s)})$$
$$= \prod_{0 \leq s, r < p} \eta^{r+s} \prod_{0 \leq s, r < p} 2i \sin \frac{r-s}{p} \pi.$$

Because $\sum_{r > s \geq 0}^{p-1} r + s = 2p((p-1)/2)^2$ we have $\prod \eta^{r+s} = 1$. Hence

$$A = i^{(p-1)p/2} 2^{p(p-1)/2} \prod_{0 \leq s < r < 0} \sin \frac{(r-s)\pi}{p}.$$

Because we know the form of $F(p)^2$ we see that

$$A^2 = p^p (-1)^{p(p-1)/2}$$

and

$$A = \pm i^{p(p-1)/2} p^{p/2}.$$

Since $\sin[(r-s)\pi/p] > 0$ for $0 \leqslant s < r \leqslant p-1$ we must have

$$\det A = i^{p(p-1)/2} p^{p/2} = p^{p/2} \det(F(p)).$$

Hence $i^{p(p-1)/2} = (-1)^{m_2} i^{m_3} (-i)^{m_4}$. Thus

$$\frac{p(p-1)}{2} \equiv 2m_2 + m_3 - m_4 \mod 4.$$

From which we have

$$m_1 - m_2 \equiv \bmod 4 \quad \text{if } p \equiv 1 \bmod 4,$$
$$m_3 - m_4 \equiv \bmod 4 \quad \text{if } p \equiv 3 \bmod 4.$$

This combines with our previous results to

$$\operatorname{Tr}(F(p)) = \begin{cases} 1 & \text{if } p \equiv \bmod 4, \\ i & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

I.4.2. McClellan and Park's Proof of Theorem I.1.2′. The proof of Theorem I.1.2′ by McClellan and Park is interesting for three reasons. First, it

is the first direct proof of Theorem I.1.2′; second, it is extremely explicity, in that it exhibits eigenvectors; third, it rests on the use of Chebyshev sets. Using Chebyshev sets is a very novel idea and we do not believe it would have occurred to many "pure" mathematicians. It is also interesting to note that McClellan and Park were ignorant of Schur's work (see [15]). We will now try to present the flavor of their proof. For complete details see [16].

We will list some basic facts about Chebyshev sets.

DEFINITION. A set of n smooth functions on an interval is a Chebyshev set for the interval if any nonzero element in the linear span of the n functions has at most $n-1$ distinct zeros.

THEOREM C.1.

$\{1, \cos t, \ldots, \cos nt\}$ *is a Chebyshev set on* $[0, \pi]$.

$\{\sin t, \ldots, \sin nt\}$ *is a Chebyshev set on* $(0, \pi)$.

THEOREM C.2. *If* $\{\varphi_1(t), \ldots, \varphi_n(t)\}$ *is a Chebyshev set on an interval and* $t_1, \ldots, t_{n+1}$ *are distinct points of the interval, then the matrix*

$$\begin{bmatrix} \phi_1(t_1) & \cdots & \phi_n(t_1) \\ \vdots & & \vdots \\ \phi_1(t_n) & \cdots & \phi_n(t_n) \end{bmatrix}$$

is nonsingular. If δ, $i = 1, \ldots, n+1$, *are all nonzero and alternate in sign, then*

$$\begin{bmatrix} \phi_1(t_1) & \cdots & \phi_n(t_1)\delta_1 \\ \vdots & & \vdots \\ \phi_1(t_{n+1}) & \cdots & \phi_n(t_{n+1})\delta_{n+1} \end{bmatrix}$$

is a nonsingular matrix.

We will now list certain elementary facts about the finite Fourier transform. Proofs can be found in [16].

DEFINITION. A function f on $\mathbf{Z}/n$ is called even if

$$f(\alpha) = f(-\alpha), \qquad \alpha \in \mathbf{Z}/n.$$

A function f on $\mathbf{Z}/n$ is called odd if

$$f(\alpha) = -f(-\alpha), \qquad \alpha \in \mathbf{Z}/n.$$

1. For $f \in L^2(\mathbf{Z}/n)$, $F^2(n)(f)(\alpha) = f(-\alpha)$.

2. Let $[x]$ denote the greatest integer less than x. Then $L^2(\mathbf{Z}/n)$ has a $v = [n/2] + 1$ dimensional subspace of even functions and an $n - v$ dimensional subspace of odd functions.

3. If f is an eigenvector of $F(n)$, then f is either an even or odd function.

4. Even eigenvectors have eigenvalues ± 1. Odd eigenvectors have eigenvalues $\pm i$.

5. If f is an even function, then $F(n)(f) + f\,(F(n)(f) - f)$ is an eigenvector of $F(n)$ with eigenvalue 1 (-1). If f is an odd function, then $iF(n)(f) -$

$f(iF(n)(f) + f)$ is an eigenvector of $F(n)$ with eigenvalue i $(-i)$.

6. If g is an even (odd) eigenvector of $F(n)$, then there exists an even (odd) function f such that $g = F(n)(f) \pm f$ $(g = iF(n) \mp f)$. We are now in a position to outline the McClellan-Park proof.

Consider the cases $N = 4m$, $4m + 1$, $4m + 2$, $4m + 3$ separately. Let m_i, $i = 1, 2, 3, 4$, be as in the statement of Theorem I.1.2′. The steps of the proof are as follows: Exhibit f_k even functions $1 \leqslant k \leqslant m_1$ such that $F(n)f_k + f_k$, $k = 1, \ldots, m_1$, are linearly independent. Since $F(n)f_k + f_k$ has eigenvalue 1, this will prove that the multiplicity of the eigenvalue 1 is greater than or equal to m_1. Exhibit f_k' even functions $1 \leqslant k \leqslant m_2$ such that $F(n)f_k' + f_k'$ are linearly independent. Since $F(n)f_k - f_k$ has eigenvalue -1, this will prove that the multiplicity of the eigenvalue -1 is greater than or equal to m_2. Similar statements hold for m_3 and m_4. Since $m_1 + m_2 + m_3 + m_4 = N$, this will prove Theorem I.1.2′.

We will indicate the method of proof by working out for the case $N = 4m$ that the multiplicity of the eigenvalue 1 is greater than or equal to $m_1 = m + 1$. Let f_α, $\alpha = 0, \ldots, 4m - 1$, be the bases of $L^2(\mathbf{Z}/N)$ where f_α is the function that takes the value 1 at $\alpha \in \mathbf{Z}/N$ and zero at all other points. Let

$$g_0 = f_0, \qquad g_m = f_{2m}, \qquad g_i = f_i + f_{N-i}, \quad i = 1, \ldots, m - 1.$$

We need to study

$$\sum_{i=0}^{m} a_i(F(N)g_i + g_i)$$

or

$$(F(N) + I)\sum_{i=0}^{m} a_i g_i = 0.$$

Since $e^{-2\pi ik/N} = e^{2\pi i(N-k)/N}$ the coefficients of $f_m, \ldots, f_{2m}$ yield $m + 1$ equations in $m + 1$ unknowns that can be written as

$$\begin{bmatrix} 1 & \cos t_1 & \cdots & \cos(m-1)t_1 & \cdots & (-1)^m \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & \cos t_m & \cdots & \cos(m-1)t_m & \cdots & (-1)^{2m-1} \\ 1 & \cos t_{m+1} & \cdots & \cos(m-1)t_{m+1} & \cdots & (1+\sqrt{N}) \end{bmatrix} \begin{bmatrix} a_0 \\ 2a_1 \\ \vdots \\ 2a_m \end{bmatrix}$$

where $t_i = [(m + i - 1)/m][\pi/2]$, $i = 1, \ldots, m + 1$. We now apply Theorem C.2 to conclude that the images of the vectors $g_0, \ldots, g_m$ are independent and so the multiplicity of the eigenvalue 1 is greater than or equal to $m + 1$ for $N = 4m$.

I.4.3. Dirichlet's Proof of Theorem I.1.2. We will now outline Dirichlet's proof of Theorem I.1.2. Complete details can be found, for instance, in Lang **[14]**.

This proof of Theorem I.1.2 rests on the following classical result about Fourier series.

THEOREM F.1. *If θ is a C' function on* [0, 1] *then*

$$\theta(x) = \sum_{m=-\infty}^{\infty} c_m e^{2\pi i m x}, \qquad 0 < x < 1,$$

and

$$\frac{\theta(0) + \theta(1)}{2} = \sum_{m=-\infty}^{\infty} c_m$$

where

$$c_m = \int_0^1 \theta(x) e^{-2\pi i m x}\, dx.$$

Now let $f(x) = e^{2\pi i x^2/n}$, $0 \leqslant x \leqslant 1$, *and let* $f_k(x) = f(x + k)$, $k = 0, \ldots, n - 1$. *Then*

$$\frac{f_k(0) + f_k(1)}{2} = \frac{e^{2\pi i k^2} + e^{2\pi i (k+1)^2/n}}{2}.$$

Hence

$$\sum_{k=0}^{n-1} \frac{f_k(0) + f_k(1)}{2} = \frac{1 + e^{2\pi i/n}}{2} + \frac{e^{2\pi i/n} + e^{2\pi i 2^2/n}}{2} + \cdots + \frac{e^{2\pi i (n-1)^2/n} + e^{2\pi i n^2/n}}{2}.$$

By reassociating the terms in the sum, we obtain

$$\sum_{k=0}^{n-1} \frac{f_k(0) + f_k(1)}{2} = \frac{1}{2} + \sum_{k=1}^{n-1} e^{2\pi i k^2/n} + \frac{1}{2} = \sqrt{n}\ \mathrm{Tr}(F(n)).$$

Let $\theta = f_0 + \cdots + f_{n-1}$. Then θ is C' on [0, 1] and so, by Theorem F.1, we have

$$\sqrt{n}\ \mathrm{Tr}(F(n)) = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{n-1} \int_0^1 f_k(x) e^{-2\pi i m x}\, dx.$$

After some elementary operations that include completing the square, we obtain

$$\sqrt{n}\ \mathrm{Tr}(F(n)) = \sum_{m=-\infty}^{\infty} e^{-\pi i n m^2/2} \int_0^n e^{2\pi i (x - nm/2)^2/n}\, dx.$$

If m is even $e^{-\pi i n m^2/2} = 1$ and if m is odd $e^{-\pi i n m^2/2} = i^{-n}$. We split the sum over even m and odd m. A computation that involves letting $m = 2r$ or $m = 2r + 1$ shows that the sums of the integrals over m even or m odd are equal to

$$I_n = \int_{-\infty}^{\infty} e^{2\pi i/n y^2}\, dy.$$

One verifies that the above improper integral converges and that $I_n = \sqrt{n}\, I_1$. From all this we otain

$$\mathrm{Tr}(F(n)) = (1 + i^{-n})/(1 + i^{-1})$$

which is another form of Theorem I.1.2.

I.4.4. LANDSBERG'S PROOF OF THEOREM I.1.2. We will now outline Landsberg's proof of Theorem I.1.2 as presented in Bellman [5]. This proof is particularly interesting in light of Theorem I.3.3 and the fact that the algebra

$$\mathbf{C}[X_1, X_2^2, X_3^3]/(X_3^6 + X_1^4X_2^2 + X_2^6)$$

is the algebra of theta functions of characteristic (0, 0) and period i. Landsberg's proof rests on the famous functional equation satisfied by the theta constants. To be more precise, the theta constants are the first order theta functions of period $t = r + is$, $r > 0$, evaluated at the origin. This is the function

$$f(t) = \sum_{n=-\infty}^{\infty} e^{-n^2t}$$

and $f(t)$ satisfies the functional equation

$$f(t) = \left(\frac{\pi}{t}\right)^{1/2} f\left(\frac{\pi^2}{t}\right).$$

Clearly $f(t)$ diverges along the line $\mathrm{Re}(t) = 0$. To find the relation between Gaussian sums and the theta constant $f(t)$, we examine $f(t)$ in a neighborhood of its line of divergence.

Let $S(p, q) = \sum_{r=0}^{q-1} e^{-\pi ir^2p/q}$, $(p, q) = 1$. Set $t = \varepsilon + \pi ip/q$ where $\varepsilon > 0$. Then

$$f\left(\varepsilon + \frac{\pi ip}{q}\right) = 1 + 2\sum_{r=1}^{q} e^{-\pi ir^2p/q}\left\{\sum_{s=0}^{\infty} e^{-(r+sq)^2\varepsilon}\right\}.$$

The function of ε in the right-hand bracket behaves like the integral

$$\int_0^{\infty} e^{-(r+sq)^2\varepsilon}\, ds = \int_r^{\infty} e^{-\omega^2\varepsilon}\, d\omega/q.$$

Now as $\varepsilon \to 0$ the above integral is asymptotic to

$$\frac{1}{q\sqrt{\varepsilon}} \int_0^{\infty} e^{-\omega^2}\, d\omega = \frac{\sqrt{\pi}}{2q\sqrt{\varepsilon}}.$$

Hence as $\varepsilon \to 0$, we have $f(\varepsilon + \pi ip/q)$ is asymptotic to $\sqrt{\pi}\, S(p, q)/q\sqrt{\varepsilon}$.

One similarly finds that $f(\pi^2/t)$ is asymptotic to $\sqrt{\pi}\, S(-q, p)/q\sqrt{\varepsilon}$ as $\varepsilon \to 0$. Using the fundamental functional equation and the asymptotics discussed above, we find

$$\frac{1}{\sqrt{q}} \sum_{r=0}^{q-1} e^{-\pi ir^2p/q} = \frac{e^{-\pi i/4}}{\sqrt{p}} \sum_{r=0}^{p-1} e^{\pi ir^2q/p}.$$

Choosing q odd and $p = 2$ yields Theorem I.1.2 for n odd.

5. The finite Heisenberg groups and the finite Fourier transform. In this section we will begin to discuss the role that nilpotent harmonic analysis plays in the theory of the finite Fourier transform. The reason for the importance of nilpotent harmonic analysis is that certain finite nilpotent groups have the finite Fourier transform built into their structure. The first indications of this are given in this section and will be looked at again in Chapter II.

Let $\mathfrak{R}$ be a commutative ring with identity and such that $2x = 0$, $x \in \mathfrak{R}$, implies $x = 0$. We define the $\mathfrak{R}$-Heisenberg group as the group of matrices

$$\begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

with $a, b, c \in \mathfrak{R}$. It is easy to see that the $\mathfrak{R}$-Heisenberg may be defined as the set

$$N(\mathfrak{R}) = \mathfrak{R} \times \mathfrak{R} \times \mathfrak{R} = \{(a, b, c) | a, b, c \in \mathfrak{R}\}$$

with multiplication given by

$$(a_1, b_1, c_1)(a_2, b_2, c_2) = (a_1 + a_2, b_1 + b_2, c_1 + c_2 + a_2 b_2).$$

The center of $N(\mathfrak{R})$, $z(N(\mathfrak{R})) = \{(0, 0, c) | c \in \mathfrak{R}\}$ and $N(\mathfrak{R})/z(N(\mathfrak{R})) \approx \mathfrak{R} \oplus \mathfrak{R}$. Clearly $z(N(\mathfrak{R}))$ is isomorphic to the $\mathfrak{R}$ as an additive group.

We will now review some of the basic facts about unitary representations of the groups $N(\mathbf{Z}/n)$.

Let $\mathbf{C}^m$ denote the m-dimensional complex vector space $(c_1, \ldots, c_m) = \mathbf{c}$. Define the usual Hermitian structure on $\mathbf{C}^m$ by

$$\langle \mathbf{c}, \mathbf{d} \rangle = \sum_{i=1}^{m} c_i \bar{d}_i$$

where bar denotes complex conjugate. Let $U(m)$ denote the group of linear transformation of $\mathbf{C}^m$ such that for $U \in U(m)$ and $\mathbf{c}, \mathbf{d} \in \mathbf{C}^m$

$$\langle \mathbf{c}, \mathbf{d} \rangle = \langle U(\mathbf{c}), U(\mathbf{d}) \rangle.$$

$U(m)$ is then called the group of unitary *transformations* of $\mathbf{C}^m$. A *unitary representation* ρ of a group G is a homomorphism of G into $U(m)$. A unitary representation ρ is called irreducible if the only subspace of $\mathbf{C}^m$ invariant under $\rho(g)$, $g \in G$, is $\mathbf{C}^m$ or 0. Also, ρ is called faithful if ρ is a monomorphism.

Let ρ_1 and ρ_2 be unitary representations of G on $\mathbf{C}^m$. We will say that ρ_1 is unitarily equivalent to ρ_2 if there exists a unitary matrix U such that

$$U^{-1} \rho_1(g) U = \rho_2(g), \quad \text{all } g \in G.$$

U is then called an intertwining operator for ρ_1 and ρ_2.

Let I be the identity matrix in $U(m)$ and let χ be a character on an abelian group A. By χI we mean the unitary representation of A defined by $(\chi I)(a) = \chi(a) I$, $a \in A$.

Let G be a group and let $z(G)$ denote the center of G.

THEOREM R.1. *Let ρ_1 and ρ_2 be unitary representations of G and assume, in addition, that ρ_1 is irreducible. Let U be an intertwining operator for ρ_1 and ρ_2. Then ρ_2 is irreducible and U is unique up to multiplication by an element of $\mathbf{C}_1^{\times}$.*

THEOREM R.2. *Let ρ be an irreducible representation of G and let $z(G)$ denote the center of G. Then z restricted to $z(G)$, $\rho/z(G) = \chi \cdot I$, where χ is a character of $z(G)$.*

We will now state two results that are specific for the groups $N(\mathbf{Z}/n)$.

These results give a picture of the faithful irreducible unitary representations of $N(\mathbf{Z}/n)$.

THEOREM R.3. *Let ρ_1 and ρ_2 be two irreducible unitary representations of $N(\mathbf{Z}/n)$ then ρ_1 and ρ_2 are unitary equivalent if and only if $\rho_1|z(N(\mathbf{Z}/n)) = \rho_2|z(N(\mathbf{Z}/n))$.*

THEOREM R.4. *Let A be a maximal abelian subgroup of $N(\mathbf{Z}/n)$ and let χ be a character on A such that $\chi|z(N(\mathbf{Z}/n))$ is a faithful character. Then inducing χ from A to $N(\mathbf{Z}/n)$ gives an irreducible unitary representation of $N(\mathbf{Z}/n)$.*

Thus, every faithful character χ of $z(N(\mathbf{Z}/n))$ extends to a faithful irreducible unitary representation of $N(\mathbf{Z}/n)$ and every faithful irreducible unitary representation of $N(\mathbf{Z}/n)$ restricts to a faithful character on $z(N(\mathbf{Z}/n))$.

In order to introduce two irreducible unitary representations ρ_1 and ρ_2 of $N(\mathbf{Z}/n)$ that have the finite Fourier transform as intertwining operator; i.e.,

$$F(n)\rho_2 F(n)^{-1} = \rho_1$$

we need to define the following matrices.

$$D_n(a) = \begin{bmatrix} e^{2\pi i a \cdot 0/n} & & 0 \\ & \ddots & \\ 0 & & e^{2\pi i(n-1)a/n} \end{bmatrix}, \qquad 0 \leqslant a < n,$$

$$S_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & & & & 1 \\ 1 & 0 & \cdots & & 0 \end{bmatrix}$$

Let $A = (a, 0, c)$ and let $\chi(0, 0, c) = e^{2\pi i c/n}$. Inducing χ from A to $N(\mathbf{Z}/n)$ gives the following irreducible unitary representation of $N(\mathbf{Z}/n) = \{(a, b, c) | a, b, c \in \mathbf{Z}/n\}$

$$\rho_1(z(N(\mathbf{Z}/n))) = \chi \cdot I,$$
$$\rho_1(a, 0, 0) = D_n(a),$$
$$\rho_1(0, b, 0) = (S_n)^b.$$

Let $B = (0, b, c)$ and let $\chi(0, 0, c) = e^{2\pi i c/n}$. Inducing χ from B to $N(\mathbf{Z}/A)$ gives the following irreducible unitary representations of $N(\mathbf{Z}/n)$

$$\rho_2(z(N(\mathbf{Z}/n))) = \chi \cdot I,$$
$$\rho_2(a, 0, 0) = (S_n)^a,$$
$$\rho_2(0, b, 0) = D_n(b).$$

It is an elementary exercise to verify that ρ_1 and ρ_2 are irreducible unitary representations of $N(\mathbf{Z}/n)$ and if

$$F(n) = \frac{1}{\sqrt{n}}(e^{2\pi i ab/n}), \qquad 0 \leqslant a,b < n,$$

that

$$F(n)\rho_2 F(n)^{-1} = \rho_1$$

or $F(n)$ intertwines ρ_1 and ρ_2. By the uniqueness of intertwining operators, $F(n)$ is essentially determined by the representations ρ_1 and ρ_2.

At this point the appearance of $F(n)$ as an intertwining operator for unitary representations of $N(\mathbf{Z}/n)$ is totally unexplained. In §II.3 we will see another way of looking at ρ_1 and ρ_2 that better explains the finite Fourier transform's role as an intertwining operator for ρ_1 and ρ_2.

6. A proof of Theorem I.3.3, nil-theta functions and theta functions. In this section we will outline a proof of Theorem I.3.3 that uses harmonic analysis on the real Heisenberg group and is completely independent of Theorems I.1.2 or I.1.2′. This proof shows the deep relation between the finite Fourier transforms, $F(n)$, $n > 0$, and the algebra of theta functions with periods 1 and $\sqrt{-1}$. A complete exposition of this material can be found in Chapter II of [1].

Let $\mathbf{R}$ denote the reals and $\mathbf{Z} \subset \mathbf{R}$, denote the integers. Let $N = N(\mathbf{R})$ be the $\mathbf{R}$-Heisenberg group and let $\Gamma = N(\mathbf{Z})$ be the $\mathbf{Z}$-Heisenberg group. Then $\Gamma \subset N$ and $\Gamma \setminus N$ is a compact manifold. If $(x, y, z) \in N$, $x, y, z \in \mathbf{R}$, then the 3-form $dx \wedge dy \wedge dz$ induces a probability measure on $\Gamma \setminus N$. We form the Hilbert space $L^2(\Gamma \setminus N)$ and define a unitary representation U of N on $L^2(\Gamma \setminus N)$ as follows: For $g \in N, f \in L^2 \in (\Gamma \setminus N)$ define

$$(U(g)f)(\Gamma h) = f(\Gamma hg), \qquad h \in N.$$

It will be convenient to consider functions $\Gamma \setminus N$ as functions on N such that $f(\gamma h) = f(h)$, $\gamma \in \Gamma$, $h \in N$. In general, if f is a function on N and $\gamma \in N$ we set

$$\mathfrak{L}_r(f)(h) = f(r^{-1}h)$$

and call $\mathfrak{L}$ the left action. For each $m \in \mathbf{Z}$, let

$$H(m) = \{f \in L^2(\Gamma \setminus N) | f(x, y, z + t) = e^{2\pi i m t} f(x, y, z)\}.$$

One verifies that $U(g)H(m) = H(m)$, $g \in N$, and that

$$L^2(\Gamma \setminus N) = \sum_{m \in \mathbf{Z}} \oplus H(m)$$

where the sum is the orthogonal sum.

We now want to better understand the spaces $H(m)$. To do this we introduce the automorphism

$$D_m : N \to N$$

given by $N_m(x, y, z) = (mx, y, mz)$. By letting $f \to f \circ D_m$ we may use D_m to induce a linear mapping of $L^2(\Gamma \setminus N)$. By a slight abuse of notation we will denote this linear mapping also by D_m. Then

$$D_m(H(1)) \subset H(m).$$

One verifies that

$$H(m) = \sum_{j=0}^{m-1} \mathfrak{L}_{(0,j/m,0)}(D_m(H(1))).$$

We will denote $\mathcal{L}_{(0,j/m,0)}(D_m(H(1)))$ by $H(m,j)$, $0 \leqslant j \leqslant m-1$. Then $U(g)H(m,j) = H(m,j)$, $g \in N$, all j. A deeper fact is that each of the spaces $H(m,j)$ is irreducible under the action of U and that $H(m,h)$ and $H(n,k)$ are unitary equivalent with respect to U if and only if $m = n$.

The notions of irreducibility and unitary equivalence are the general Hilbert space analogues of those introduced in the previous section. We will say more about the structure of $H(1)$ in §II, where we consider the Weil-Brezin map.

We next observe that

$$J(x,y,z) = (-y, x, z - xy)$$

is an automorphism of N such that $J(\Gamma) = \Gamma$ and such that $J(H(m)) = H(m)$. Hence J induces a unitary operator on $H(m)$ which we call J_m. The action of J_1 on $H(1)$ corresponds to the real Fourier transform in a sense that will be discussed in Chapter II.

In [1] we showed that J enables us to define a first order differential operator $D(J)$ on $\Gamma \setminus N$ such that

$$D(J)f = D(J)(J(f)), \qquad f \in C^\infty(\Gamma \setminus N).$$

Let

$$\Theta(m) = \{f \in C^\infty \cap H(m) | D(J)f = 0\}, \qquad m > 0.$$

We will now outline the main properties of the subspaces $\Theta(m)$ and $\Theta = \Sigma_{m>0} \oplus \Theta(m)$. (Notice, since $f \in \Theta(m)$ and $g \in \Theta(n)$ are both C^∞ functions on $\Gamma \setminus N$ their product fg is a C^∞ function on $\Gamma \setminus N$.) First, Θ is an algebra and $\Theta(m)\Theta(n) \subset \Theta(m+n)$. Second, $J(\Theta(m)) = \Theta(m)$. Third, $\dim \Theta(m) = m$. Finally, Θ has no zero divisors.

The representation theory can be used to prove that

$$\Theta(m) = \sum_{j=0}^{m-1} \oplus \Theta(m,j)$$

where $\Theta(m,j) = \Theta(m) \cap H(m,j) = \mathcal{L}_{(0,j/m,0)}(D_m(\Theta(1)))$ and $\Theta(1)$ has basis

$$\varphi(x,y,z) = e^{2\pi i z} \sum_{l \in \mathbf{Z}} e^{-\pi(y+l)^2} e^{2\pi i l x}.$$

The relationship between the algebra Θ and J and the space $\mathcal{Q} = \Sigma_{n>0} \oplus L^2(\mathbf{Z}/n)$ and the finite Fourier transform extend to $\mathcal{Q}$ by $F = \Sigma_{n>0} \oplus F(n)$ is given by the following theorem.

THEOREM I.6.1. *There exists a unitary operator* $V\colon \Theta \to \mathcal{Q}$ *satisfying*
(1) $V(\Theta(n)) = L^2(\mathbf{Z}/n)$,
(2) $F = VJV^{-1}$.

It follows that $\mathcal{Q}$ can be given the algebra structure of Θ and since $J(f_1 f_2) = J(f_1)J(f_2)$, $f_1, f_2 \in \Theta$, we have

$$F(g_1 g_2) = F(g_1)F(g_2), \qquad g_1, g_2 \in \mathcal{Q}.$$

It is also proved in [4] that Θ is isomorphic to

$$\mathbf{C}[X_1, X_2^2, X_3^3] / (X_3^6 + X_1^4 X_2^2 + X_2^6)$$

and so Theorem I.3.1 follows from Theorem I.6.1.

The proof of Theorem I.6.1 proceeds in the following way. We need to find a basis of $\Theta(m)$ such that the matrix of Jm with respect to this basis is

$$J_m = \frac{1}{\sqrt{m}} e^{2\pi i\alpha\beta/m}, \qquad 0 \leqslant \alpha,\beta < m.$$

We do this as follows. Let

$$\varphi_m(x, y, z) = e^{2\pi iz} \sum_{l \in \mathbf{Z}} e^{-\pi m(y+l)^2} e^{2\pi ilx}.$$

Then $D_m(\varphi_m) \in \Theta(m, 0)$ and we can form the basis

$$\varphi_{mj} = L_{(0,j/m,0)}(D_m(\varphi_m)), \qquad 0 \leqslant j < m,$$

of $\Theta(m)$. In [1] we show that φ_{mj}, $0 \leqslant j < m$, is the required basis.

We can also verify this result using the ideas of §I.5. First, we define a representation U_1 of the $\mathbf{Z}/m$-Heisenberg group $N(\mathbf{Z}/m)$ on $\Theta(m)$ as follows. For $f \in \Theta(m)$ and $a, b, c \in \mathbf{Z}$ let

$$U_1(m)(a, 0, 0)f = L_{(a/m,0,0)}f,$$
$$U_1(m)(0, b, 0)f = L_{(0,b/m,0)}f,$$
$$U_1(m)(0, 0, c)f = e^{2\pi i(c/m)}f.$$

It is not hard to see that with respect to the basis φ_{mj}, defined above, that $U_1(m)$ is a unitary representation of $N(\mathbf{Z}/m)$ on $\Theta(m)$ and that the matrix $U_1(m) = \rho_1$, ρ_1 as defined in the previous section. Also one verifies that

$$J^{-1}U_1(m)J = \rho_2.$$

Since $J(m)$ intertwines ρ_1 and ρ_2 it follows that $J(m) = cF(m)$ where $|c| = 1$. One then verifies that $c = 1$ and we have our assertion.

Chapter II. The Cooley-Tukey Algorithm and the Weil-Brezin Map

1. The Cooley-Tukey algorithm. Currently the most popular algorithm for computing the finite Fourier transform is called the Cooley-Tukey algorithm. The history of this algorithm has been set forth in an interesting article by Cooley et al. [9] and the original paper is Cooley and Tukey [10]. It is our intention in this section to analyze in some detail the basic construction upon which the Cooley-Tukey algorithm rests. This will enable us to relate the Cooley-Tukey algorithm to the Weil-Brezin map (see [19], [7]) and the proof of the Plancherel theorem for the reals as given in Chapter 1 of [3].

Because it has been so important in the theory of numerical computations and because it is so brief, we will begin by reproducing the few paragraphs in Cooley and Tukey [10] that–aside from induction–set forth the idea of the Cooley-Tukey algorithm.

"Consider the problem of calculating the complex Fourier series

$$X(j) = \sum_{k=0}^{N-1} A(k) W^{jk}, \qquad j = 0, \ldots, N-1, \tag{1}$$

where the given Fourier coefficients $A(k)$ are complex and W is the principal Nth root of unity

$$W = e^{2\pi i/N}. \tag{2}$$

A straightforward calculation using (1) would require N^2 operations where 'operation' means, as it will throughout this note, a complex multiplication followed by a complex addition.

The algorithm described here iterates on the array of given complex Fourier amplitudes and yields the result in less than $2N \log_2 N$ operations without requiring more data storage than is required for the given array A. To derive the algorithm, suppose N is composite, i.e., $N = r_1 r_2$. Then let the indices in (1) be expressed

$$\begin{aligned} j &= j_1 r_1 + j_0, \quad j_0 = 0, 1, \ldots, r_1-1, \quad j_1 = 0, 1, \ldots, r_2-1, \\ k &= k_1 r_2 + k_0, \quad k_0 = 0, 1, \ldots, r_2-1, \quad k_1 = 0, 1, \ldots, r_1-1. \end{aligned} \tag{3}$$

Then, one can write

$$X(j_1, j_0) = \sum_{k_0} \sum_{k_1} A(k_1, k_0) W^{jk_1 r_2} W^{jk_0} \tag{4}$$

since

$$W^{jk_1 r_2} = W^{j_0 k_1 r_2}. \tag{5}$$

The inner sum, over k_1, depends only on j_0 and k_0 and can be defined as a new array,

$$A_1(j_0, k_0) = \sum_{k_1} A(k_1, k_0) W^{j_0 k_1 r_2}. \tag{6}$$

The result can then be written

$$X(j_1, j_0) = \sum_{k_0} A_1(j_0, k_0) W^{(j_1 r_1 + j_0) k_0}. \tag{7}$$

There are N elements in the array A_1, each requiring r_1 operations, giving a total of Nr_1 operations to obtain A_1. Similarly, it takes Nr_2 operations to calculate X from A_1. Therefore, this two-step algorithm, given by (6) and (7), requires a total of

$$T = N(r_1 + r_2)$$

operations."

Let us now formalize the steps of the Cooley-Tukey algorithm. We let $\mathcal{F}(\,)$ denote the complex functions whose domain is the set in the parens and

$\mathcal{F}(X, Y)$ denote the mappings from X to Y. Then the Cooley-Tukey algorithm rests on the commutativity of the following diagram:

$$\begin{array}{ccccc}
\mathcal{F}(\mathbf{Z}/r_1r_2) & \xrightarrow{C} & \mathcal{F}(\mathbf{Z}/r_1 \times \mathbf{Z}/r_2) \approx \mathcal{F}(\mathbf{Z}/r_1, \mathcal{F}(\mathbf{Z}/r_2)) & & \\
 & & & & \downarrow 1 \times F(r_2) \\
 & & \mathcal{F}(\mathbf{Z}/r_1 \times \widehat{\mathbf{Z}/r_2}) \approx \mathcal{F}(\mathbf{Z}/r_1, \mathcal{F}(\mathbf{Z}/r_2)) & & \\
F(r_1r_2) \downarrow & & \downarrow M & & \\
 & & \mathcal{F}(\widehat{\mathbf{Z}/r_2} \times \mathbf{Z}/r_1) \approx \mathcal{F}(\widehat{\mathbf{Z}/r_2}, \mathcal{F}(\mathbf{Z}/r_1)) & & \\
 & & & & \downarrow 1 \times F(r_1) \\
\mathcal{F}(\widehat{\mathbf{Z}/r_1r_2}) & \xleftarrow{D^{-1}} & \mathcal{F}(\widehat{\mathbf{Z}/r_2} \times \widehat{\mathbf{Z}/r_1}) \approx \mathcal{F}(\widehat{\mathbf{Z}/r_2}, \mathcal{F}(\widehat{\mathbf{Z}/r_1})) &
\end{array}$$

where the "hat" denotes the dual group or the group of characters and where we must still define the various mappings of the above diagram.

DEFINITIONS. THE MAPPING $\approx$. Let $f(x, y) \in \mathcal{F}(X \times Y)$. Then for each fixed $x_0 \in X$, $f(x_0, y) \in \mathcal{F}(Y)$ and so $f(x, y)$ determines an element of $\mathcal{F}(X, \mathcal{F}(Y))$. It is obvious that this correspondence is 1-1.

THE MAPPING C. Between sets $\mathbf{Z}/r_1r_2$ and $\mathbf{Z}/r_1 \times \mathbf{Z}/r_2$ define the following homeomorphism C^*. (Notice: C^* is not a group homomorphism.) For $0 \leqslant k < r_1r_2$ let $k = k_2 + k_1r_2$, where $0 \leqslant k_2 < r_2$, $0 \leqslant k_1 < r_1$. Define

$$C^*(k) = (k_1, k_2) \in \mathbf{Z}/r_1 \times \mathbf{Z}/r_2.$$

For $f \in \mathcal{F}(\mathbf{Z}/r_1r_2)$ define $C(f) = f \circ (C^*)^{-1} \in \mathcal{F}(\mathbf{Z}/r_1 \times \mathbf{Z}/r_2)$.

THE MAPPING D. Between the sets $\widehat{\mathbf{Z}/r_1r_2}$ and $\widehat{\mathbf{Z}/r_2} \times \widehat{\mathbf{Z}/r_1}$ define the homeomorphism D^* as follows: For $0 \leqslant k < r_1r_2$ let $k = k_2r_1 + k_1$, where $0 \leqslant k_2 < r_2$, $0 \leqslant k_1 < r_1$. Define

$$D^*(k) = (k_2, k_1) \in \widehat{\mathbf{Z}/r_2} \times \widehat{\mathbf{Z}/r_1}.$$

For $f \in \mathcal{F}(\mathbf{Z}/r_1r_2)$ define $D(f) = f \circ (D^*)^{-1} \in \mathcal{F}(\widehat{\mathbf{Z}/r_2} \times \widehat{\mathbf{Z}/r_1})$.

THE MAPPINGS $1 \times F(r_2)$ AND $1 \times F(r_2)$. An element of $\mathcal{F}(\mathbf{Z}/r_1, \mathcal{F}(\mathbf{Z}/r_2))$ determines an r_1-tuple of elements of $\mathcal{F}(\mathbf{Z}/r_2)$. The mapping $1 \times F(r_2)$ denotes applying the Fourier transform $F(r_2)$ to each of these r_1-tuple of elements of $\mathcal{F}(\mathbf{Z}/r_2)$. Define $1 \times F(r_1)$ similarly.

THE MAPPING M. For $0 \leqslant a < r_1$ and $0 \leqslant b < r_2$ define

$$M(F)(b, a) = e^{2\pi iab/r_1r_2}F(a, b), \qquad F \in \mathcal{F}(\mathbf{Z}/r_1 \times \mathbf{Z}/r_2).$$

For those people who believe that the Fourier transform is related to the groups $\mathbf{Z}/n$ and $\widehat{\mathbf{Z}/n}$ as presented in modern texts in Harmonic Analysis, the mappings C, D, and M involved in the Cooley-Tukey algorithm seem at best formal and at worst arbitrary. We will show in the next section that if we use the representation theory of the finite Heisenberg group, then the mappings C, D and M are natural.

2. The finite Heisenberg groups and the Cooley-Tukey algorithm. Before going into the details of this section, we will, as promised, present some of the material in §I.5 in a slightly different language. This language has the advantage of showing the deep inter-relation of the finite Fourier transform $F(n)$, the dual pairing of $\mathbf{Z}/n$ and its dual group $\widehat{\mathbf{Z}/n}$, and the finite Heisenberg group $H(n)$.

Let $\mathbf{C}^{\times}(n)$ denote the multiplicative group of complex numbers $e^{2\pi ik/n}$, $k = 0, \ldots, n-1$. Let $G(n)$ as a set be $\mathbf{Z}/n \times \widehat{\mathbf{Z}/n} \times \mathbf{C}^{\times}(n)$ and for $(a_i, \hat{b}_i, c_i) \in G(n)$, $i = 1, 2$, define multiplication by

$$(a_1, \hat{b}_1, c_1)(a_2, \hat{b}_2, c_2) = (a_1 + a_2, \hat{b}_1 + \hat{b}_2, c_1 \cdot c_2 \cdot \langle a_1, \hat{b}_2 \rangle)$$

where $\langle\ ,\ \rangle$ denotes the dual pairing of $\mathbf{Z}/n$ and $\widehat{\mathbf{Z}/n}$ to $\mathbf{C}^{\times}(n)$. Then $G(n)$ is a group with this law of composition. We claim that $G(n)$ is isomorphic to $H(n)$. Let

$$\alpha: \mathbf{Z}/n \to \mathbf{C}^{\times}(n)$$

be defined by $\alpha(k) = e^{2\pi ik/n}$. It follows that α is an isomorphism between the additive group $\mathbf{Z}/n$ and the multiplicative group $\mathbf{C}^{\times}(n)$.

We now define β: $\mathbf{Z}/n \to \mathbf{Z}/n$ as follows: For $k \in \mathbf{Z}/n$ define

$$\langle l, \beta(k) \rangle = e^{2\pi ilk/n}, \quad \text{all } l \in \mathbf{Z}/n.$$

Now define γ: $H(n) \to G(n)$ by

$$(a_1, a_2, a_3) \to (a_1, \beta(a_2), \alpha(a_3)).$$

It is an elementary computation to verify that γ is an isomorphism. This shows that the finite Heisenberg group is built from the group structures on $\mathbf{Z}/n$ and $\mathbf{Z}/n$ combined with the dual pairing of $\mathbf{Z}/n$ and $\mathbf{Z}/n$.

In this general setting, the finite Fourier transform is the isometry $F(n)$: $L^2(\mathbf{Z}/n) \to L^2(\mathbf{Z}/n)$ defined by

$$(F(n)(f))(\hat{a}) = \frac{1}{\sqrt{n}} \sum_{a \in \mathbf{Z}/n} f(a)\langle a, \hat{a} \rangle, \qquad \hat{a} \in \mathbf{Z}/n.$$

Define the action of $\mathbf{Z}/n$ on $L^2(\mathbf{Z}/n)$ as follows: For $a, b \in \mathbf{Z}/n$ define

$$(T(a)f)(b) = f(b + a).$$

Similarly define $T(\hat{a})$, for $\hat{a}, \hat{b} \in \widehat{\mathbf{Z}/n}$, of $\hat{f} \in L^2(\widehat{\mathbf{Z}/n})$ by

$$(T(\hat{a})\hat{f})(\hat{b}) = \hat{f}(\hat{b} + \hat{a}).$$

Let $G_F(n)$ be the group of linear transformations of $L^2(\mathbf{Z}/n)$ generated by $T(a)$, $a \in \mathbf{Z}/n$, and $F^{-1}(n)T(\hat{a})F(n)$, $\hat{a} \in \mathbf{Z}/n$. We wish to now obtain a matrix representation of the group $G_F(n)$ in order to understand what the group $G_F(n)$ is really like. To do this, let f_α, $\alpha = 0, \ldots, n-1$, be the basis of $L^2(\mathbf{Z}/n)$, where f_α takes the value 1 at $\alpha \in \mathbf{Z}/n$ and 0 at all other points. Clearly $T(1)(f_\alpha) = f_{\alpha+1}$, and so, relative to the basis $f_0, \ldots, f_{n-1}$, $T(1)$ has the

matrix representation

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ & 0 & 1 & \ddots & 0 \\ 0 & & & & 1 \\ 1 & 0 & & \cdots & 0 \end{bmatrix}.$$

Since $(T(\hat{1})F(n)f_\alpha)(\hat{a}) = \langle \alpha, \hat{1}\rangle\langle \alpha, \hat{a}\rangle/\sqrt{n}$, we have

$$(F(n)^{-1}T(\hat{1})F(n))f_\alpha = \langle a, \hat{1}\rangle.$$

Thus, using the isomorphism γ defined above, we have $F(n)^{-1}T(\hat{1})F(n)(f_\alpha)$ has the matrix representation

$$\begin{bmatrix} e^{2\pi i 0/n} & & 0 \\ & \ddots & \\ 0 & & e^{2\pi i(n-1)/n} \end{bmatrix}.$$

This proves that $G_F(n)$ has the matrix representation ρ_2 of $N(\mathbf{Z}/n)$. This argument shows the deep relation between the group $N(\mathbf{Z}/n)$ and the finite Fourier transform $F(n)$. It also proves that

$$F(n)\rho_2 F(n)^{-1} = \rho_1$$

as it is easy to see that $\rho_1(N(\mathbf{Z}/n))$ is the same as the matrix group generated by $T(\hat{a})$, $\hat{a} \in \mathbf{Z}/n$, and $F(n)T(a)F(n)^{-1}$, $a \in \mathbf{Z}/n$.

We now proceed to the task of characterizing the mappings C, D and M discussed in the previous section.

Consider the $\mathbf{Z}/n$-Heisenberg group, $N(\mathbf{Z}/n)$, $n = r_1 r_2$, $r_1 > 1$ and $r_2 > 1$ where

$$N(\mathbf{Z}/n) = \{(a, b, c) | a, b, c \in \mathbf{Z}/n\}.$$

Let $\Gamma(r_2, r_1)$ and $\Gamma(r_1, r_2)$ contained in $N(\mathbf{Z}/n)$ be defined by

$$\Gamma(r_2, r_1) = \{(a'r_2, b'r_1, 0) | a' \in \mathbf{Z}/r_1, b' \in \mathbf{Z}/r_2\},$$
$$\Gamma(r_1, r_2) = \{(b'r_1, a'r_2, 0) | a' \in \mathbf{Z}/r_1, b' \in \mathbf{Z}/r_2\}.$$

An elementary computation shows that $\Gamma(r_2, r_1)$ and $\Gamma(r_1, r_2)$ are subgroups of $N(\mathbf{Z}/n)$. Let Γ be a subgroup of $N(\mathbf{Z}/n)$. Consider the homogeneous space $\Gamma \setminus N(\mathbf{Z}/n)$ and give this finite set the measure where each point has measure one. Form $L^2(\Gamma \setminus N(\mathbf{Z}/n))$. Since $N(\mathbf{Z}/n)$ acts on $\Gamma \setminus N(\mathbf{Z}/n)$ by

$$R(g)(\Gamma n) = \Gamma n g, \qquad n, g \in N(\mathbf{Z}/n),$$

$R(g)$, $g \in N(\mathbf{Z}/n)$, defines a unitary representation R of $N(\mathbf{Z}/n)$ on $L^2(\Gamma \setminus N(\mathbf{Z}/n))$ by

$$(R(g)(F))(\Gamma n) = F(\Gamma n g), \qquad F \in L^2(\Gamma \setminus N(\mathbf{Z}/n)).$$

It will often be convenient to view functions on $\Gamma \setminus N(\mathbf{Z}/n)$ as functions F on $N(\mathbf{Z}/n)$ such that

$$F(\gamma g) = F(g), \qquad \gamma \in \Gamma, g \in N(\mathbf{Z}/n).$$

Let χ be the character on the center $z(N(\mathbf{Z}/n))$ given by

$$\chi(0, 0, c) = e^{2\pi ic/n}.$$

Then $-\chi$ is the character given by

$$-\chi(0, 0, c) = e^{-2\pi ic/n}.$$

Let $\mathfrak{F}(\chi)$ denote the functions $F(a, b, c)$ on $N(\mathbf{Z}/n)$ such that

$$F(a, b, c + d) = e^{2\pi id/n}F(a, b, c).$$

Define $\mathfrak{F}(-\chi)$ analogously. Let

$$\mathfrak{F}(\chi, \Gamma(r_2, r_1)) = \mathfrak{F}(\chi) \cap L^2(\Gamma(r_2, r_1) \setminus N(\mathbf{Z}/n))$$

and

$$\mathfrak{F}(-\chi, \Gamma(r_1, r_2)) = \mathfrak{F}(-\chi) \cap L^2(\Gamma(r_1, r_2) \setminus N(\mathbf{Z}/n)).$$

It is easily verified that

$$R(\mathfrak{F}(\chi, \Gamma(r_2, r_1))) = \mathfrak{F}(\chi, \Gamma(r_2, r_1)),$$
$$R(\mathfrak{F}(-\chi, \Gamma(r_1, r_2))) = \mathfrak{F}(-\chi, \Gamma(r_1, r_2)).$$

Let $R^*(\chi)$ denote the restriction of R to $\mathfrak{F}(\chi, \Gamma(r_2, r_1))$ and $R^*(-\chi)$ denote the restriction of R to $\mathfrak{F}(-\chi, \Gamma(r_1, r_2))$.

THEOREM II.2.1. *$R^*(\chi)$ is an irreducible unitary representation of $N(\mathbf{Z}/n)$ that is unitarily equivalent to ρ_1 or ρ_2.*

PROOF. $R^*(\chi)(0, 0, d)$ is $e^{2\pi id}$ by our definitions of $\mathfrak{F}(\chi)$. Hence by Theorems R.1 and R.3 the proof of our assertion reduces to computing the dimension of $\mathfrak{F}(\chi, \Gamma(r_2, r_1))$. But the dimension of $\mathfrak{F}(\chi, \Gamma(r_2, r_1))$ is easily seen to be the same as the dimension of $\mathfrak{F}(\mathbf{Z}/r_1 \times \mathbf{Z}/r_2)$ which is r_1r_2 or n. Hence $R^*(\chi)$ is irreducible.

(A similar argument shows that $R^*(-\chi)$ is irreducible.)

Since $R^*(\chi)$, is unitarily equivalent to ρ_2, there exists a unitary operator

$$W: L^2(\mathbf{Z}/n) \to \mathfrak{F}(\chi, \Gamma(r_2, r_1))$$

such that $W^{-1}R^*(\chi)W = \rho_2$. Recall that W is unique up to multiplication by a complex number of absolute value 1.

We will now build W from $1 \times F(r_2) \circ C$, where C and $1 \times F(r_2)$ are as defined earlier in this chapter. For $f \in L^2(\mathbf{Z}/n)$ define

$$W(f) = F(x, y, t) \in \mathfrak{F}(\mathbf{Z}/n \times \mathbf{Z}/n \times \mathbf{Z}/n)$$

by

$$W(f) = e^{2\pi it/n} \sum_{0 \le j < r_1} f(jr_2 + x)e^{2\pi ijr_2y/n}.$$

W is the analogue of the Weil-Brezin map as defined in [19] and [7]. We will now verify that $W(f) \in \mathfrak{F}(\chi, \Gamma(r_2, r_1))$. It is clear that $W(f) \in \mathfrak{F}(\chi)$. Hence it remains to verify that

$$W(f)((ar_2, br_1, 0)(x, y, t)) = W(f)(x, y, t).$$

Now $(ar_2, br_1, 0)(x, y, t) = (ar_2 + x, br_1 + y, t + ar_2y)$. Thus the left side

above equals

$$e^{2\pi i(t+ar_2y)/n}\sum f(jr_2+ar_2+x)e^{2\pi ijr_2(y+br_1)/n}$$

$$= e^{2\pi it/n}\sum f((y+a)r_2+x)e^{2\pi ijr_2(y+br_1)/n}e^{2\pi i(ar_2y)/n}$$

$$= e^{2\pi it/n}\sum f((j+a)r_2+x)e^{2\pi i(j+a)r_2y/n} = W(f)(x,y,t),$$

and we have shown that $W(f) \in \mathcal{F}(\chi, \Gamma(r_2, r_1))$ or

$$W\colon L^2(\mathbf{Z}/n) \to \mathcal{F}(\chi, \Gamma(r_2, r_1)).$$

To relate W to $1 \times F(r_2) \circ C$, we note that $W(f)$ restricted to the set $S = \{(x, y, 0) | 0 \leqslant x < r, 0 \leqslant y < r_1\}$ equals $(1 \times F(r_2) \circ C)f$. Further, since $W(f) \in \mathcal{F}(\chi, \Gamma(r_2, r_1))$, knowing $W(f)$ on the set S uniquely determines $W(f)$.

It is straightforward from the discussion above and the properties of the mappings $1 \times F(r_2)$ and C to conclude that W is a unitary operator.

It remains to verify that

$$W^{-1}R^*(\chi)W = \rho_2.$$

But this is a formal calculation that the interested reader may easily verify. This shows that $1 \times F(r_2) \circ C$ is essentially an intertwining operator between two irreducible unitary representations of $N(\mathbf{Z}/n)$.

We come next to the mapping M of the Cooley-Tukey algorithm. To explain M one has to introduce a bit more of the structure of $N(\mathbf{Z}/n)$. To be precise, we must introduce a particular automorphism K of the group $N(\mathbf{Z}/n)$. For $(x, y, t) \in N(\mathbf{Z}/n)$ let

$$K(x, y, t) = (x, y, -t + xy).$$

It is a straightforward computation to verify that K is an automorphism of $N(\mathbf{Z}/n)$ and that K^2 is the identity automorphism.

Consider the following general situation. Let B be a group, G a subgroup of B and A an automorphism of B with $A(G_1) = G_2$. Let $f \in \mathcal{F}(B)$ be such that

$$f(gb) = f(b), \qquad b \in B, g \in G_1.$$

If $g \in G_1$, $A^{-1}g \in G_2$, and so if $f(gb) = f(b)$, $g \in G_1$, we have

$$f(A^{-1}(gb)) = f(A^{-1}(g)A^{-1}(b)) = f(A^{-1}(b)).$$

Hence if $f^* = f \circ A^{-1}$ we have

$$f^*(g_2b) = f^*(b), \qquad g_2 \in G_2 \text{ and } b \in B.$$

Now apply this to the special case of the functions $W(f)$ and the automorphism K above. Because $K = K^{-1}$ we have $W(f) \circ K$ is invariant under $\Gamma(r_1, r_2)$ and so is in $\mathcal{F}(-\chi, r_1, r_2)$. Explicitly

$$W(f)(K(x, y, t)) = W(f)(y, x, -t + xy)$$

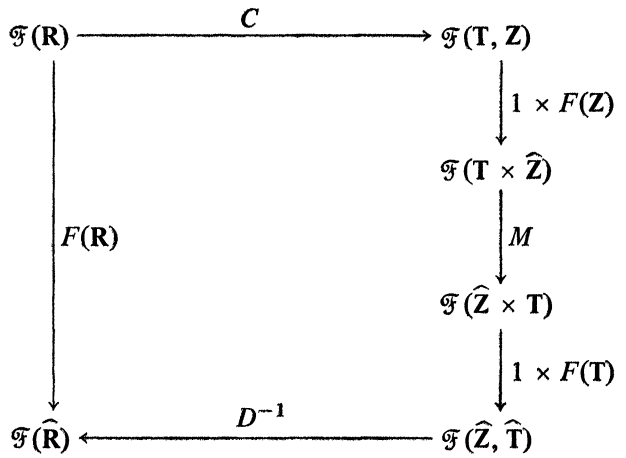
$$= \sum_{0 \leqslant a < r} f(ar_2 + y)e^{2\pi iar_2x/n}e^{-2\pi it/n}e^{2\pi ixy/n}.$$

But the mapping $W(f) \circ K$ on $(x, y, 0)$ is the same as applying M to $W(f)(x, y, 0)$. This supplies us with the group theoretic interpretation of M that we sought.

Now $R^*(-\chi)$ an irreducible unitary representation of $N(\mathbf{Z}/n)$ on $\mathfrak{F}(-\chi, r_1, r_2)$. One can show that $D^{-1} \circ 1 \times F(r_1)(x, y, 0)$ determines an intertwining operator between $R^*(-\chi)$ and an irreducible unitary representation of $N(\mathbf{Z}/n)$ on $L^2(\widehat{\mathbf{Z}/r_1 r_2})$. Since the discussion is similar to that given above for $1 \times F(r_2) \circ C$, we will not go any further into the specific details.

3. The Plancherel theorem for the reals and the Cooley-Tukey algorithm. We will now show the general nature of the Cooley-Tukey algorithm by showing how it can be used to prove the Plancherel theorem for the reals. Just as for the Cooley-Tukey algorithm, the forthcoming proof of the Plancherel theorem has an interpretation in terms of nilpotent harmonic analysis. We will not present this material because a full discussion would be quite long. For the interested reader the material in Chapter I of [3] or Chapter I, §5, of [1] can be modified along the lines of the material in the previous section to obtain all the group theoretic ramifications of our method of proof.

Let $\mathbf{R}$ denote the reals, $\mathbf{Z}$ the integers and $\mathbf{T}$ the circle group or $\mathbf{R}/\mathbf{Z}$, and let $F(\ )$ denote the Fourier transform of the group in the parens. The diagram of the Cooley-Tukey algorithm becomes in this setting the diagram below.



Recall that as groups $\hat{\mathbf{R}}$ is isomorphic to $\mathbf{R}$, $\hat{Z}$ is isomorphic to $\mathbf{T}$ and $\hat{\mathbf{T}}$ is isomorphic to $\mathbf{Z}$.

Define C^*: $\mathbf{R} \to [0, 1) \times \mathbf{Z}$, where we identify $\mathbf{T}$ with $[0, 1)$, as follows: If $x \in \mathbf{R}$, $x = y + n$, $0 \leqslant y < 1$, $n \in \mathbf{Z}$, let

$$C^*(x) = (y, n).$$

For $f \in \mathfrak{F}(\mathbf{R})$ define $C(f) = f \circ C^{*-1}$. For $F(y, n) \in \mathfrak{F}(\mathbf{T}, \mathbf{Z})$ define

$$1 \times F(\mathbf{Z})(F(y, n)) = \sum_{n=-\infty}^{\infty} F(y, n) e^{2\pi i n \xi}, \qquad 0 \leqslant \xi < 1.$$

M is now interchange of ξ and y and multiplication by $e^{2\pi i \xi y}$. This gives

$$((1 \times F(\mathbf{T})) \circ M \circ (1 \times F(\mathbf{Z})) \circ C)f = \int_0^1 \sum_{n=-\infty}^{\infty} F(y, n) e^{2\pi i n \xi} e^{2\pi i \xi y} e^{2\pi i m y}\, dy.$$

For the time being, let us proceed formally and interchange integration and

summation to obtain

$$(1 \times F(\mathbf{T})) \circ M \circ (1 \times F(\mathbf{Z})) \circ C(f) = \sum_{n=-\infty}^{\infty} \int_0^1 F(y, n) e^{2\pi i(n+y)(m+\xi)}\, dy.$$

Applying D^{-1} we obtain

$$F(\mathbf{R})(f)(2\pi s) = \int_{-\infty}^{\infty} f(x) e^{2\pi i x s}\, ds$$

where $x = n + y$, $s = m + \xi$.

We will now see how all this works rigorously.

Consider $L^2(\mathbf{R})$ with Haar measure dx and $\mathbf{T} \times \mathbf{Z}$ with Haar measure such that the measure of $\mathbf{T} \times 0$ is one. It is easily seen that C: $L^2(\mathbf{R}) \to L^2(\mathbf{T} \times \mathbf{Z})$ is a unitary operator; i.e., a norm preserving surjection. Assume that $F(\mathbf{T})$: $L^2(\mathbf{T}) \to L^2(\mathbf{Z})$ and $F(\mathbf{Z})$: $L^2(\mathbf{Z}) \to L^2(\mathbf{T})$ are unitary operators. Since M consists of multiplying each value of a function by a number of absolute value 1, it is trivial to verify that M is a unitary operator. This shows that

$$D^{-1} \cdot 1 \times F(\mathbf{T}) \cdot M \cdot 1 \times F(\mathbf{Z}) \cdot C\colon L^2(\mathbf{R}) \to L^2(\hat{\mathbf{R}})$$

is a unitary operator and shows that $F(\mathbf{R})$ is a unitary operator.

We are left with the task of interpreting the interchange of integration and summation used in deriving the integral formula for $F(\mathbf{R})$. We will close this chapter with a discussion of this process.

We may view

$$G(y, \xi) = M \cdot 1 \times F(\mathbf{Z}) \cdot C(f) \in L^2(\mathbf{T} \times \mathbf{T}).$$

As such, $G(y, \xi)$ has a Fourier expansion

$$G(y, \xi) = \sum_{a,b \in \mathbf{Z}} B_{ab} e^{-2\pi i(a\xi + by)}$$

where convergence is in the L^2 norm. Now

$$\int_0^1 G(y, \xi) e^{2\pi i m y}\, dy = 1 \times F(\mathbf{T})(G(y, \xi)).$$

But the integral on the left is easily seen to be

$$\int_{m=-\infty}^{\infty} B_{am} e^{-2\pi i a \xi} = \hat{f}(2\pi(\xi + m)).$$

We next note that if g_m and $g \in L^2(\mathbf{T})$ with $\lim_{n\to\infty} g_n = g$ in $L^2(\mathbf{T})$, then the mth Fourier coefficient of g_n converges to the mth Fourier coefficient of g. This shows that if $f \in L^2(\mathbf{R})$, then

$$\hat{f}(2\pi s) = \lim_{n\to\infty} \int_{-n}^{n} f(x) e^{2\pi i s t}\, ds$$

where the limit is in the L^2 norm.

Chapter III. Algebraic Complexity and the Finite Fourier Transform

1. Basic ideas in algebraic complexity. S. Winograd's work on algebraic complexity is very interesting at both the practical and the theoretical levels. He has in [20], [21] produced algorithms for the finite Fourier transform that

are more efficient than Cooley-Tukey. At the theoretical level, he has also succeeded in [22] in defining the concept of *essential multiplications or divisions*, or more briefly, *essential* m/d. This concept is important because experience has shown that algorithms that minimize the essential m/d, or minimal algorithms, possess interesting algebraic structures. We begin with a formal definition of an algorithm that will permit us to define the concept of essential m/d.

Let G be a field, called the field of constants, and let $F = G(x_1, \ldots, x_n)$ be the purely transcendental field extension of G obtained by adjoining the indeterminants $x_1, \ldots, x_n$ to G. We use $\Omega = \{\omega_i | i = 1, 2, 3, 4\}$ to denote the field operations of addition, subtraction, multiplication and division, respectively. For $f_1, f_2 \in F$ we will use $\omega_i(f_1, f_2)$ to denote the result of applying the binary operation ω_i to f_1 and f_2 with the convention that

$$\omega_4(f_1, f_2) = f_1/f_2.$$

Let $B \subset F$ be the *given objects* for our algorithms. Usually, $B = G \cup \{x_1, \ldots, x_n\}$. But often, in theoretical discussions, other given objects will play important roles.

DEFINITION. We will define an N-step algorithm α over (F, B) inductively.

Step 1. Choose either an element of B or choose $\omega(1) \in \Omega$ and an ordered pair $(a(1), b(1))$ from B. Require that $O_\alpha(1) = \omega(1)(a(1), b(1))$ be defined and call $O_\alpha(1)$ the output of the first step of the algorithm.

Step 2. Choose either an element of B or choose $\omega(2) \in \Omega$ and an ordered pair $(a(2), b(2))$ from $B \cup O_\alpha(1)$. Require that $O_\alpha(2) = \omega(2)(a(2), b(2))$ be defined and call $O_\alpha(2)$ the output of Step 2 of the algorithm.

Assume the first k steps of α have been defined.

Step $k + 1$. Choose either an element of B or choose $\omega(k + 1) \in \Omega$ and an ordered pair $(a(k + 1), b(k + 1))$ from $B \cup O_\alpha(1) \cup \cdots \cup O_\alpha(k)$. Require that

$$O_\alpha(k + 1) = \omega(k + 1)(a(k + 1), b(k + 1))$$

be defined, and call $O_\alpha(k + 1)$ the output of the $k + 1$ step of the algorithm.

If α has N steps, we will call it an N-step algorithm. We call $O_\alpha(k)$, $1 \leqslant k \leqslant N$, the output function of the algorithm α.

Two algorithms α, β over (F, B) will be said to be equivalent if

$$O_\alpha(k) = O_\beta(k), \qquad 1 \leqslant k \leqslant N.$$

The k step of an algorithm α is called an m/d step if $\omega(k)$ is multiplication or division; i.e., if $\omega(k) = \omega_3$ or ω_4. Clearly equivalent algorithms need not have the same number of m/d steps. (For instance, $x + x = 2x$ and $2 \cdot x = 2x$.) This may serve to motivate the following definition.

DEFINITION. A step k for an algorithm α is called m/d essential if $O(k)$ is not in the G-linear span of $B \cup O(1) \cup \cdots \cup O(k - 1)$.

DEFINITION. Let $f_1, \ldots, f_s \in F$. We will say that the N-step algorithm α over (F, B) computes $f_1, \ldots, f_s$ if for each f_i, $1 \leqslant i \leqslant s$, there is an integer $k(i)$, $1 \leqslant k(i) \leqslant N$, such that $O_\alpha(k(i)) = f_i$.

It is obvious that $f_1, \ldots, f_s \in F$ can be computed by an algorithm over (F, B) if and only if $f_1, \ldots, f_s$ are in the field generated by B.

DEFINITION. We will say that α is a *minimal algorithm* for computing $f_1, \ldots, f_k$ if, among all algorithms over (F, B), α has the minimum number of essential m/d.

The *m/d number* for computing $f_1, \ldots, f_k$ over (F, B) is the number of essential m/d steps in a minimal algorithm for computing $f_1, \ldots, f_k$.

Let α be an N-step algorithm over (F, B) with output function $O_\alpha(k)$, $1 \leqslant k \leqslant N$. Let $s\colon B \to F$ be a mapping. Then, under certain circumstances, s determines an N-step algorithm $s(\alpha)$ over $(F, s(B))$. We will now describe this.

DEFINITION. Let $B \subset F$ and let $[B]$ be the subring of F generated by B and (B) be the subfield generated by B. A mapping $s\colon B \to F$ is called a *basis for substitution* if the following are satisfied:

(a) There is a ring homomorphism $s^*\colon [B] \to F$ such that $s^*|B = s$.

(*Note*. Since B generates $[B]$ as a ring, if s^* exists, it is unique.)

(b) For $b_1, b_2 \in [B]$, if $s^*(b_2) \neq 0$, $s^*(b_1)/s^*(b_2)$ is well defined. (Hence s^* can be extended to as much of (B) as possible.)

Let α be an N-step algorithm over (F, B) and $s\colon B \to F$ a basis of substitution. Let the k steps of α be $\omega(k)$, $(a(k), b(k))$. Then $s(\alpha)$ is defined and has k step $\omega(k)$, $(s^*(a(k)), s^*(b(k)))$ or $s(O_\alpha(k))$ if $O_\alpha(k) \in B$ provided $s^*(b(k)) \neq 0$ whenever $\omega(k) = \omega_4$. If $s(\alpha)$ is defined then $s(\alpha)$ is an N-step algorithm over $(F, s(B))$ and $O_{s(\alpha)}(k) = s^*(O_\alpha(k))$.

DEFINITION. Let α be an N-step algorithm over (F, B) and β an M-step algorithm over (F, B'). Let $\alpha \circ \beta$ be the $N + M$-step algorithm over $(F, B \cup B')$ whose k step, $1 \leqslant k \leqslant M$, is the k step of β and whose k step, $M + 1 \leqslant k \leqslant M + N$, is the $k - M$ step of α.

Let α be an N-step algorithm over (F, B) that computes $f_1, \ldots, f_r$ and let $a(k)$ or $b(k)$, $1 \leqslant k \leqslant N$, be in the subset $b_1, \ldots, b_s$ of B. Let β be an (F, B') algorithm that computes $b_1, \ldots, b_s$. Then $\alpha \circ \beta$ is an (F, B') algorithm that computes $f_1, \ldots, f_r$.

In particular, if s is a basis of substitution such that $s(\alpha)$ is an $(F, s(B))$ algorithm and β is an (F, B') algorithm computing $s^*(a_\alpha(k))$ or $s^*(b_\alpha(k))$ for $a_\alpha(k)$ or $b_\alpha(k) \in B$, we have $s(\alpha) \circ \beta$ over (F, B') computes $s(\alpha)$.

2. Bilinear algorithms for the finite Fourier transform. S. Winograd has recently devised algorithms for computing the finite Fourier transform that work much better than the Cooley-Tukey algorithm. They are also of theoretical interest because they are based on expressing the finite Fourier transform $F(p)$, p a prime, in terms of the complex group algebra of the multiplicative group $\mathbf{Z}/p^x$. We will denote this group algebra by $\mathbf{C}(\mathbf{Z}/p^x)$.

Let us begin by writing Winograd's algorithm for $p = 7$. Let

$$A_j = \sum_{k=0}^{6} e^{2\pi ijk/7} a_k, \qquad j = 0, \ldots, 6.$$

We will present the algorithm as a sequence of additions, then multiplications and then additions.

$$S_1 = a_1 + a_6, \quad S_2 = a_1 - a_6, \quad S_3 = a_4 + a_3, \quad S_4 = a_4 - a_3,$$
$$S_5 = a_2 - a_5, \quad S_6 = a_2 + a_5, \quad S_7 = S_1 + S_3, \quad S_8 = S_7 + S_6,$$
$$S_9 = S_8 + a_0, \quad S_{10} = S_1 - S_3, \quad S_{11} = S_3 - S_5, \quad S_{12} = S_5 - S_1,$$
$$S_{13} = S_2 + S_4, \quad S_{14} = S_{13} + S_6, \quad S_{15} = S_2 - S_4, \quad S_{16} = S_4 - S_6,$$
$$S_{17} = S_6 - S_2,$$

Let $u = 2\pi i/7$.

$$m_1 = \left(\frac{\cos u + \cos 2u + \cos 3u}{3} - 1\right) S_8,$$
$$m_2 = \left(\frac{2\cos u - \cos 2u - \cos 3u}{3}\right) S_{10},$$
$$m_3 = \left(\frac{\cos u - 2\cos 2u + \cos 3u}{3}\right) S_{11},$$
$$m_4 = \left(\frac{\cos u + \cos 2u - 2\cos 3u}{3}\right) S_{12},$$
$$m_5 = i\left(\frac{\sin u + \sin 2u - \sin 3u}{3}\right) S_{14},$$
$$m_6 = i\left(\frac{2\sin u - \sin 2u + \sin 3u}{3}\right) S_{15},$$
$$m_7 = i\left(\frac{\sin u - 2\sin 2u - \sin 3u}{3}\right) S_{16},$$
$$m_8 = i\left(\frac{\sin u + \sin 2u + 2\sin 3u}{3}\right) S_{17}.$$

$$S_{18} = S_9 + m_1, \quad S_{19} = S_{18} + m_2, \quad S_{20} = S_{19} + m_3, \quad S_{21} = S_{18} - m_2,$$
$$S_{22} = S_{21} - m_4, \quad S_{23} = S_{18} - m_3, \quad S_{24} = S_{23} + m_4, \quad S_{25} = m_5 + m_6,$$
$$S_{26} = S_{25} + m_7, \quad S_{27} = m_5 - m_6, \quad S_{28} = S_{27} - m_8, \quad S_{29} = m_5 - m_7, .$$
$$S_{30} = S_{29} + m_8, \quad S_{31} = S_{20} + S_{26}, \quad S_{32} = S_{20} - S_{26}, \quad S_{33} = S_{22} + S_{28},$$
$$S_{34} = S_{22} - S_{28}, \quad S_{35} = S_{24} + S_{30}, \quad S_{36} = S_{24} - S_{30}$$
$$A_0 = S_9, \quad A_1 = S_{31}, \quad A_2 = S_{33}, \quad A_3 = S_{36},$$
$$A_4 = S_{35}, \quad A_5 = S_{34}, \quad A_6 = S_{32}.$$

This algorithm requires 8 multiplications instead of $8 \cdot 3 = 24$ that Cooley-Tukey requires for $F(8)$.

We now will describe some of the theoretical considerations upon which the above algorithm was built.

Let $\omega = e^{2\pi i/p}$, p a prime, and $\omega^{jk} = e^{2\pi ijk/p}$, $0 \leqslant j, k < p$, and

$$A_j = \sum_{k=0}^{p-1} e^{2\pi ijk/p} a_k = \sum_{k=0}^{p-1} \omega^{jk} a_k. \tag{1}$$

Since $\omega^{jk} = 1$ for j or $k = 0$ we have $A_0 = a_0 + \cdots + a_{p-1}$ and

$$A_j - a_0 = A_j^*, \qquad 1 \leqslant j \leqslant p - 1, \tag{2}$$

where

$$A_j^* = \sum_{k=1}^{p-1} \omega^{jk} a_k, \qquad 1 \leqslant j \leqslant p-1. \tag{3}$$

Let $\{m_1, \ldots, m_{p-1}\}$ be the elements of $(\mathbf{Z}/p)^x$ ordered so that $m_i \cdot m_j = m_k$, where $k = i \cdot j \bmod p$. Then

$$\left(\sum_{j=1}^{p-1} \omega^j m_j\right)\left(\sum a_k (m_k)^{-1}\right) = \sum_l \sum_k \omega^{lk} a_k m_l \tag{4}$$

because $m_j (m_k)^{-1} = m_l$ implies $j = l \cdot k \bmod p$. Since $\Sigma A_l^* m_l$ is the right side of (4), computing the finite Fourier transform is the same as computing a product in $\mathbf{C}(\mathbf{Z}/p^x)$. Because $\mathbf{Z}/p^x$ is a cyclic group of order $p-1$, we can obtain another method for computing the terms A_l^*, $l = 1, \ldots, p-1$. Let $m \in \mathbf{Z}/p^x$ be such that $m^{p-1} = 1 \in \mathbf{Z}/p^x$; i.e., m is a generator of the multiplicative cyclic group $\mathbf{Z}/p^x$. Then $m = m_\alpha$ for some $1 < \alpha < p-1$. Hence

$$\left(\sum_{j=1}^{p-1} \omega^j m_j\right) = \sum_{j=1}^{p-1} \omega^{\alpha^j m^j} \quad (\text{here } k\, m^j = (m)^j)$$

and

$$\left(\sum y_k m_k^{-1}\right) = \sum y_{\pi(\alpha^j)} m^j$$

where $\pi(\alpha^j) \cdot \alpha^j = 1 \bmod p$. Hence

$$\sum A_{\alpha^j}^* m^j = \left(\sum \omega^{\alpha^j m^j}\right)\left(\sum y_{\pi(\alpha^j)} m^j\right).$$

But, multiplication on the right side is the same as multiplying the expressions as polynomials in m and reducing modulo $(m^{p-1}-1)$. This directly relates the finite Fourier transform to the group algebra $\mathbf{C}(\mathbf{Z}/p-1)$.

We may also present the above discussion in matrix language as follows: Consider the matrix equation

$$\begin{bmatrix} A_1^* \\ \vdots \\ A_{p-1}^* \end{bmatrix} = (\omega^{jk}) \begin{bmatrix} a_1 \\ \vdots \\ a_{p-1} \end{bmatrix},$$

$$A^* = \Omega a.$$

The first row of the square matrix on the right is $\omega^1, \ldots, \omega^{p-1}$. Thus there is a permutation of columns of this square matrix so that the first row of the resulting matrix is $\omega^\alpha, \ldots, \omega^{\alpha^{p-1}}$. This permutation can be achieved by multiplying the matrix (ω^{ij}) on the right by a matrix P. Noting that $P^{-1} = P^t$, where the superscript t denotes the transpose, we have

$$A^* = \Omega P P^t a.$$

Then the first column of Ω is $(\omega^1, \ldots, \omega^{p-1})^t$. Forming

$$P^t A^* = P^t \Omega P P^t a$$

an elementary computation shows that

$$P^t\Omega P = \begin{bmatrix} \omega^{\alpha^1} & \cdots & \omega^{\alpha^{p-1}} \\ \omega^{\alpha^2} & \cdots & \omega^{\alpha^{p-1}\omega^{\alpha}} \\ & \cdots & \\ \omega^{\alpha^{p-1}} & \omega^{\alpha^1} & \cdots \omega^{\alpha^{p-2}} \end{bmatrix}.$$

Thus, if we let

$$Y = P^tA^* = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{p-1} \end{bmatrix}; \qquad P^ta = \begin{bmatrix} y_1 \\ \vdots \\ y_{p-1} \end{bmatrix} = y$$

we have

$$Y = P^t\Omega Py$$

which is called the cyclic convolution of $(Y_1, \ldots, Y_{p-1})$ and $(y_1, \ldots, y_{p-1})$. It is now a straightforward computation to verify that

$$Y_1 + Y_2u + \cdots + Y_{p-1}u^{p-2} = (\omega^{\alpha^1} + \omega^{\alpha^2}u + \cdots + \omega^{\alpha^{p-1}}u^{p-2}) \cdot (y_1 + y_{p-1}u + \cdots + y_2u^{p-2}) \mod(y^{p-1} - 1).$$

The above discussion shows that the m/d number of $F(p)$, can be bounded above by the m/d number of the right side of the above equation.

Let us now formalize the problem to which the above discussion has led us. Let

$$R(z) = \sum_{i=0}^{a} x_iz^i \quad \text{and} \quad S(z) = \sum_{i=0}^{b} y_iz^i$$

be two polynomials with indeterminates as coefficients and let

$$T(z) = R(z) \cdot S(z).$$

The $a + b + 1$ coefficients of T are a system of bilinear forms which we will denote by $\tilde{T}$. Let P be a polynomial over G of degree n and let

$$T_p(z) = R(z) \cdot S(z) \mod P.$$

Let $\tilde{T}_p$ denote the n coefficients of T. Then $\tilde{T}_p$ is a system of bilinear forms. Let $B = G \cup \{x_1, \ldots, x_a, y_1, \ldots, y_b\}$. Our problem becomes to compute the m/d number of $\tilde{T}$ and $\tilde{T}_p$.

We will take up this problem in the next two sections. In the final section of this paper we will return and discuss how the above problem relates to the m/d number of $F(p)$, p a prime.

3. General results on bilinear algorithms. In this section, we will prove three general theorems, that, at the present state of the art, are of fundamental importance in the theory of bilinear algorithms.

Let G be a fixed infinite field and let $x_1, \ldots, x_n, y_1, \ldots, y_m$ be indetermi-

nants. Let $F = G(x_1, \ldots, x_n, y_1, \ldots, y_m) = G(x, y)$ and let

$$L_{ij} = \sum_{k=1}^{n} a_{ijk} x_k, \qquad 1 \leqslant j \leqslant m, 1 \leqslant i \leqslant t,$$

be a set of linear forms and let

$$A(x)y = (L_{ij}) \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

We call $A(x)y$ a system of bilinear forms. We wish to study the m/d number of $A(x)y$.

Notice that the columns (rows) of all possible matrices $A(x)$ form a vector space C (R) over G. We define the G-column (G-row) rank of $A(x)$ as the dimension of the vector subspace of C (R) spanned by the columns (rows) of $A(x)$.

In this section $F = G(x, y)$ and $B = G \cup \{x_1, \ldots, x_n, y_1, \ldots, y_m\}$.

THEOREM III.3.1. *Let $A(x)y$ be a system of bilinear forms. If $A(x)$ has G-row rank s then this system of bilinear forms has m/d number greater than or equal to s.*

This theorem is actually much weaker than what is known to be true. Since it is not harder to prove the more general result, we will prove it.

THEOREM III.3.1′. *Let $f_1, \ldots, f_l \in G(x, y)$ and let α be an N-step minimal algorithm for computing $f_1, \ldots, f_l$. Then the m/d number of $f_1, \ldots, f_l$ equals the dimension of the vector space W_α defined as follows:*

Let $W_\alpha^* \subset G(x, y)$ be the vector subspace spanned by $O_\alpha(1), \ldots, O_\alpha(N)$ and let L be the subspace of elements of the form $\sum g_i x_i + \sum h_j y_j + k$, $g, h, k \in G$. Define $W_\alpha = W_\alpha^* \oplus L/L$.

PROOF. Assume that $f_l = O_\alpha(N)$. (A relabelling can always achieve this.) Then let α' be the $N - 1$ step algorithm consisting of the first $N - 1$ steps of α. Then α' is a minimal algorithm for $f_1, \ldots, f_{l-1}, a_\alpha(N), b_\alpha(N)$.

We will now prove Theorem III.3.1′ by induction on N. Since a 1-step algorithm that has no essential m/d computes an element of L, we have proven the theorem for $N = 1$.

By induction the theorem is true for α' and $f_1, \ldots, f_{l-1}, a_\alpha(N), b_\alpha(N)$. Now if the N step of α is not an essential m/d step, α and α' have the same number of essential m/d steps. But, clearly, $W_{\alpha'}^* = W_\alpha^*$ and the theorem is true.

If the N step of α is an essential m/d step $W_\alpha \neq W_{\alpha'}$ or else α is not minimal. This again proves our result.

Theorem III.3.1′ implies Theorem III.3.1 once we observe that the row rank of $A(x)$ is the dimension of the vector space V spanned by $\sum a_{ijk} x_k y_j$ in $G(x, y)$. But $V \subset W_\alpha^*$ and $V \cap L = 0$. This proves dim $W_\alpha \geqslant$ dim V.

Before going on to Theorems III.3.2 and III.3.3 let us pause to give an application of Theorem III.3.1 to the problem posed in §III.2. Consider the system of bilinear forms that are the coefficients of the product of the two

polynomials

$$\sum_{i=0}^{a} x_i z^i \quad \text{and} \quad \sum_{j=0}^{b} y_j z^j.$$

This may be written in the form $A(x)y$, where $A(x)$ has the special form (assume $a \geqslant b$)

$$A(x) = \begin{bmatrix} x_0 & 0 & & \cdots & 0 \\ x_1 & x_0 & 0 & \cdots & \\ \vdots & & & & \\ x_a & & & & \\ 0 & x_a & & \cdots & x_0 \\ \vdots & & & & \\ 0 & & & & x_a \end{bmatrix}$$

where $A(x)$ has b columns and $a + b + 1$ rows and

$$y = \begin{bmatrix} y_0 \\ \vdots \\ y_b \end{bmatrix}.$$

It is easy to compute that the row rank of our special $A(x)$ is $a + b + 1$. Thus by Theorem III.3.1 the m/d number of $A(x)y$ is greater than or equal to $a + b + 1$. Later we will see that it is actually equal to $a + b + 1$.

THEOREM III.3.2. *Let $A(x)y$ be a system of bilinear forms. If $A(x)$ has G column rank s, then the m/d number of $A(x)y$ is greater than or equal to s.*

Again it is no harder to prove a slight generalization of this and we will do so. This is Theorem 1 of [22].

Consider $G(x_1, \ldots, x_n) = G(x)$. Let ϕ be a $t \times m$ matrix over $G(x)$. We shall use $\phi_1, \ldots, \phi_m$ to denote the columns of Φ. (*Note.* $A(x)$ is a $t \times m$ matrix over F.)

THEOREM III.3.2′. *Let α be an algorithm computing ϕy over $(G(x, y), B)$ where $B = G(x) \cup \{y_1, \ldots, y_m\}$. If there are s vectors in $\{\phi_1, \ldots, \phi_m\}$ such that no nontrivial linear combination of them with coefficients in G lies in the t dimensional vector space G^t, then α has at least s m/d steps of the following form: $\omega_3(a(k), b(k))$ with $a(k)$ and $b(k) \notin G$ or $\omega_4(a(k), b(k))$ with $b(k) \notin G$.*

To relate Theorem III.3.2 and Theorem III.3.2′, we must first show that if $A(x)y$ satisfies the hypothesis of Theorem III.3.2 then $A(x)y$ satisfies the hypothesis of Theorem III.3.2′. But a linear combination over G of the columns of $A(x)$ is in G^t if and only if it is the 0 vector. This shows that our first requirement is satisfied.

To see that the conclusion of Theorem III.3.2′ implies the conclusion of Theorem III.3.2 merely note that the s steps guaranteed by Theorem III.3.2′ are essential m/d for α.

Thus it remains only to prove Theorem III.3.2′. To simplify the language of this proof and only here, we have called those m/d not excluded by Theorem III.3.2′ essential.

We prove the theorem by induction on s. Clearly an algorithm with no essential m/d steps can only compute elements of the form $\sum g_j y_j + f$, $g_j \in G, f \in G(x)$. But if $s = 1$ there exists a matrix coefficient of ϕ that is not in G and so the algorithm must compute.

$$\sum h_j y_j, \quad \text{some } h_j \notin G.$$

Hence we must have at least one essential m/d in the algorithm.

Suppose the assertion holds for $a = N$. Assume Φ is such that at least $N + 1$ of the vectors $\{\phi_1, \ldots, \phi_n\}$ have no linear combination over G which is in G^t. Let α be a minimal algorithm computing ϕy and let k be the first integer such that an essential m/d step of α occurs at step k. Then, either

$$O_\alpha(k) = \left(\sum g_i y_i + f\right) \cdot \left(\sum h_i y_i + f'\right)$$

or

$$O_\alpha(k) = \frac{\sum g_i y_i + f}{\sum h_i y_i + f'}$$

for $g_i, h_i \in G$, and $f, f' \in G(x)$. Furthermore, we may assume the labelling has been done so that one of the $h_i \neq 0$, for otherwise the k step of α would not be an essential m/d step.

Clearly $s(y_n) = -f - \sum h_i y_i$, s the identity on $y_1 \cup \cdots \cup y_{n-1} \cup G(x)$. $G(x)$ is a basis of substitution. However $s(\alpha)$ may not be an algorithm. Since there are only a finite number of divisions in any algorithm the substitution $s(\alpha)$ can fail to be an algorithm only when s^* applied to some finite set $\{r_1, \ldots, r_m\} \subset G(x)[y]$ is zero. Choose $g \in G$ so that

$$s^*(r_j) + g \neq 0, \qquad j \in 1, \ldots, M.$$

This substitution yields an algorithm that computes $\phi' y'$ where $\phi'_j = \phi_j - h_i \phi_n$, $j = 1, 2, \ldots, n-1$ and $y' = (y_1, \ldots, y_{n-1})^t$. The number of essential m/d steps in α' is at least one less than in α. This is because the image of the k step in α' is not an essential m/d and the algorithm β computing $g - f' - \sum h_i y_i$ has no essential m/d. But there are at least N vectors in $\{\phi'_1, \ldots, \phi'_{n-1}\}$ such that no nontrivial G-linear combination is in G^t. Hence by induction α' has at least N essential m/d and so α has at least $N + 1$ essential m/d.

It turns out that it is much easier to study minimal algorithms that use only essential multiplications. The following theorem indicates why this is so. This is contained in the proof of Lemma 2 in [22].

THEOREM III.3.3. *Let $A(x)y$ be a system of bilinear forms and let α be an algorithm that uses the minimal number s of essential multiplicatons. Then there exist $2s$ linear forms $L_i(x, y)$, $L'_i(x, y)$, $i = 1, \ldots, s$, such that $A(x)y = Um$ where U is a matrix over G and m is the column matrix*

$$\begin{bmatrix} L_1 L_1' \\ \vdots \\ L_s L_s' \end{bmatrix}.$$

We call this a presentation theorem because it shows that no matter how complicated the original α was we can produce a minimal algorithm for computing $A(x)y$ of the following form: Compute the linear forms L_i, L_i', $i = 1, \ldots, s$, multiply, using essential multiplications, the linear forms to form the quadratic forms $L_i \cdot L_i'$, $i = 1, \ldots, s$, form linear combinations of the quadratic forms $L_i \cdot L_i'$, $i = 1, \ldots, s$, to obtain the elements $A(x)y$.

Notice the algorithm presented at the beginning of §II.2 for computing $F(7)$ was of the above form. We will call algorithms of the above form quadratic algorithms.

PROOF. Let $k_1, \ldots, k_s$ be the steps of α where the essential multiplications occur.

If m is a step of the algorithm α then

$$O_\alpha(m) = L_m(0) + L_m(x) + L_m(y) + L_m(x^2) + L_m(y^2) + L_m(xy) + \ldots$$

where $L_m(0)$ is a constant and $L_m(\ )$ is a form in the type of term in the bracket. We will denote $O_\alpha(k_i)$ by $L_i(0) + \ldots$. Then we can compute the system of bilinear forms $A(x)y$ by linear combinations over G of $O_\alpha(k_i)$, $i = 1, \ldots, s$. Hence there is a linear combination of the $L_i(xy)$, $i = 1, \ldots, s$, terms that equals each bilinear form in $A(x)y$.

It is crucial to our argument to observe that we may modify the algorithm to obtain a new algorithm α' without introducing any new essential m so that $L_i(0)$ is always zero. This is because $L_i(0) \in G$ and subtracting by it is not an essential m. We will henceforth assume that $\alpha = \alpha'$ or that the desired modification has been made.

Now $O_\alpha(k_i) = O_\alpha(l)O_\alpha(m)$ where l and m are steps of the algorithm. Note that we may form $U_l = L_l(x) + L_l(y)$ and $U_m = L_m(x) + L_m(y)$ without any essential m. Further $U_l U_m$ and $O_\alpha(l)O_\alpha(m)$ have the same quadratic terms. Since forming linear combinations preserves degree, it follows that we may replace the terms $O_\alpha(l)$ and $O_l(m)$ in our algorithm by U_l and U_m, respectively, and still compute the bilinear forms $A(x)y$. This proves our theorem.

4. Some minimal algorithms. In §III.3 we established some results that enabled us to put lower bounds on m/d numbers. In this section we will see how to use the Chinese Remainder Theorem to produce algorithms. We will also prove that in certain cases we can actually compute m/d numbers.

One version of the Chinese Remainder Theorem goes as follows. Consider the polynomial ring $G[z]$ over a field G and let $P_1, \ldots, P_k \in G[z]$ be such that P_i and P_j are relatively prime for $i \neq j$. Then the proof of the Chinese Remainder Theorem assures the existence of polynomials Q_i, $i = 1, \ldots, k$, such that

$$Q_i \equiv \delta_{ij} \mod P_j$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The usual statements then stress the

following: Given $B_1, \ldots, B_k$ in $G[z]$ there exists a B such that

$$B \equiv B_i \mod P_i.$$

Indeed

$$B = \sum B_i Q_i$$

will do.

But it also follows that if

$$B \equiv B_i \mod P_i$$

then

$$B \equiv \sum B_i Q_i \mod P \quad \text{where } P = \prod P_i.$$

It is this last assertion that we will need to construct algorithms. We will now give two applications of this idea.

Let $R(z) = \sum_{i=0}^{a} x_i z^i$ and $S(z) = \sum_{j=0}^{b} y_j z^j$. Using the notation of §III.2 we wish to compute $\tilde{T}$ or the coefficients of the polynomial $T(z) = R(z) \cdot S(z)$. We will use the Chinese Remainder Theorem to produce an algorithm that does this in $a + b + 1$ essential m/d. Combining this with the results in §III.3 we will have proven that the m/d number of $\tilde{T}$ is $a + b + 1$.

Choose $\alpha_0, \ldots, \alpha_{a+b}$ distinct elements of G and let

$$Q = \prod_{i=0}^{a+b} (z - \alpha_i).$$

Then Q is a polynomial of degree $a + b + 1$ and so $T(z)$ may be identified with $T(z) \bmod Q$.

Now let $Q_i = (z - \alpha_i)$. Then Q_i and Q_j are relatively prime for $i \neq j$. We observe that if

$$g_i = \prod_{j \neq i} (\alpha_i - \alpha_j),$$

then

$$G_i = g_i^{-1} \prod_{j \neq i} (z - \alpha_j)$$

is such that

$$G_i \equiv \delta_{ij} \mod Q_j.$$

Hence

$$\begin{aligned} T(z) = T(z) \mod Q &= \Big(\sum G_i (R \cdot S \mod Q_i) \Big) \mod Q \\ &= \sum_{i=0}^{a+b} G_i R(\alpha_i) \cdot S(\alpha_i). \end{aligned}$$

The above equation shows that $T(z)$ can be computed in the $a + b + 1$ essential multiplications $R(\alpha_i) \cdot S(\alpha_i)$.

A second method for computing $\tilde{T}$ starts by choosing $a + b$ elements of G, $\beta_1, \ldots, \beta_{a+b}$ and uses the identity

$$R(z)\cdot S(z) = R(z)\cdot S(z) \mod \prod_{i=1}^{a+b}(z-\beta_i) + x_a y_b \prod_{i=1}^{a+b}(z-\beta_i).$$

As before $R(z)\cdot S(z) \bmod \prod_{i=1}^{a+b}(z-\beta_i)$ is computed by the Chinese Remainder Theorem using $a+b$ multiplications and $x_a\cdot y_b$ is the $a+b+1$ essential multiplication.

We will say that the second method uses $\alpha_0 = \infty$ in the first method.

Now if $P = z^n + \sum_{i=0}^{n-1} g_i z^i$, since reducing mod P involves no essential m/d, we can compute $\tilde{T}_p$ if $a \geqslant n$ and $b \geqslant n$ in $2n-1$ multiplications.

Let $R(z)$ and $S(z)$ be of degree $n-1$ and let $P = \prod_{i=1}^{k} P_i$ where the P_i are pairwise relatively prime and $P_i = \bar{P}_i^{l_i}$ where $\bar{P}_i$ is irreducible. By the Chinese Remainder Theorem there exist $Q_i \in G[z]$, $i = 1, \ldots, k$, such that

$$Q_i \equiv \delta_{ij} \mod P_i$$

and

$$\begin{aligned} T(z) \mod P &\equiv \Big(\sum Q_i (R\cdot S \bmod P_i)\Big) \mod P \\ &\equiv \Big(\sum Q_i (R\cdot S) \mod P_i\Big) \mod P. \end{aligned}$$

Since multiplying by Q_i and reducing mod P involve no essential multiplications, we have that all the essential multiplications occur in computing $R\cdot S$ mod P_i. If the degree of $P_i = n_i$, then $\sum_{i=1}^{k} n_i = n$ and by the above discussion we can compute $R\cdot S$ mod P_i in $2n_i - 1$ essential multiplications. Then our algorithm takes $\sum_{i=1}^{k}(2n_i - 1) = 2n - k$ essential multiplications to compute $T(z)$ mod P or $\tilde{T}_p$ in the notation of §III.2.

We will now show that all minimal bilinear algorithms for computing $\tilde{T}$ are almost the same as the two algorithms we discussed above.

THEOREM III.4.1. *Any bilinear algorithm for computing $\tilde{T}$ in $a+b+1$ essential multiplications involves computing the $a+b+1$ bilinear forms*

$$(gR(\alpha_i))\cdot(hS(\alpha_i)),$$

where $g \neq 0$, $h \neq 0 \in G$, $\alpha_0, \ldots, \alpha_{a+b} \in G \cup \infty$ and $\alpha_i \neq \alpha_j$ for $i \neq j$. (Assume $a \geqslant b$.)

PROOF. Let $\tilde{T}$ be the set of bilinear forms defined by

$$\begin{bmatrix} \chi_0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ \chi_b & & & \chi_0 \\ \vdots & & & \vdots \\ \chi_a & & & \chi_b \\ 0 & & & \\ \vdots & & & \vdots \\ 0 & \cdots & & 0\ \chi_a \end{bmatrix} \begin{pmatrix} y_0 \\ y_b \end{pmatrix} = X_y$$

By our presentation theorem, we have that there exist $m_0, \ldots, m_{a+b}$ bilinear forms where

$$m_i = \left(\sum a_i x_i + \sum b_j y_j\right)\left(\sum a_i' x_i + \sum b_j' y_j\right)$$

and an $(a + b + 1) \times (a + b + 1)$ matrix U over G such that

$$Xy = U\begin{bmatrix} m_0 \\ \vdots \\ m_{a+b} \end{bmatrix}.$$

The essence of this theorem is to prove that each

$$\begin{aligned} m_i &= \left(g\sum_{i=0}^{a} x_i\alpha^i\right)\left(h\sum_{j=0}^{b} y_j\alpha^j\right) \\ &= k\left\{\left(\sum_{i=0}^{a} x_i\alpha^i\right)y_0 + \left(\sum_{i=0}^{a} x_i\alpha^{i+1}\right)y_1 + \cdots + \left(\sum_{i=0}^{a} x_i\alpha^{i+b}\right)y_b\right\} \end{aligned}$$

or

$$m_i = gx_a y_b. \tag{0}$$

As we discussed in §III.3, all the rows of X are linearly independent over G. Hence the set of bilinear forms Xy span an $a + b + 1$ dimensional space. Hence the set of bilinear forms Um span an $a + b + 1$ dimensional subspace of the space of bilinear forms. This implies that U is nonsingular and so we may let $W = U^{-1}$ and write

$$WXy = m. \tag{1}$$

Let $(w_0^i, \ldots, w_{a+b}^i)$ be the ith row of W. Then substitution in (1) yields

$$m_i = \left(\sum_{j=0}^{a} w_j^i x_j, \ldots, \sum_{j=0}^{a} w_{j+b}^i x_j\right)\begin{bmatrix} y_0 \\ \vdots \\ y_b \end{bmatrix}. \tag{2}$$

This implies that the bilinear form on the right side of (2) can be computed in 1 essential multiplication. By Theorem III.3.2 of §III.3 the column rank on the right side must be 1, or all the forms $\Sigma_{j=0}^{a} w_{j+k}^i x_j$, $k = 0, \ldots, b$, are all G-multiples of one. This implies that the matrix

$$\begin{bmatrix} w_0^i & w_1^i & \cdots & w_a^i \\ w_1^i & & \cdots & w_{a+1}^i \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ w_b^i & & \cdots & w_{a+b}^i \end{bmatrix} \tag{3}$$

has rank 1.

We claim this can happen only under two circumstances: either $w_0^i = 0, \ldots, w_{a+b-1}^i = 0$, $w_{a+b}^i \neq 0$; or there exists α_i (which may be zero) such

that

$$w_j^i = \alpha_i^j w_0^i, \qquad j = 0, \ldots, a+b \quad \text{where } \alpha_i^j = (\alpha_i)^j$$

(when $0^0 = 1$).

We may verify this assertion as follows: We have two cases to consider.

Case 1. $w_0^i = 0$. Then since (3) has rank 1,

$$w_0^i w_2^i - (w_1^i)^2 = 0 \quad \text{or} \quad w_1^i = 0 \tag{4}$$

and

$$w_1^i w_3^i - (w_2^i)^2 = 0 \quad \text{or} \quad w_2^i = 0. \tag{5}$$

We may proceed by induction to verify that $w_k^i = 0$, $0 \leqslant k \leqslant a+b-1$. Since rank of (3) is 1, $w_{a+b}^i \neq 0$.

Case 2. $w_0 \neq 0$. Let $w_1^i = k_1 w_0^i$ and $w_2^i = k_2 w_0^i$. By (4)

$$k_2 = k_1^2$$

and by (5)

$$k_1 k_3 = k_1^4 \quad \text{or} \quad k_3 = k_1^3.$$

We may proceed by induction to verify that if

$$w_l^i = k_l w_0^i.$$

Then $k_l = k_1^l$. We let $\alpha_i = k_1$.

Since W is nonsingular at most 1 row can be of the first kind and for two rows with $w_0^i \neq 0$ and $w_0^j \neq 0$ we must have $\alpha_i \neq \alpha_j$, $i \neq j$. Thus (2) becomes

$$w_0^i \left(\sum_{j=0}^{a} \alpha_i^j x_j, \sum_{j=0}^{a} \alpha_i^{j+1} x_j, \ldots, \sum_{j=0}^{a} \alpha_i^{j+b} x_j \right) \begin{bmatrix} y_0 \\ \vdots \\ y_b \end{bmatrix}$$

or m_i has the form (0). This proves the theorem.

$$W = \begin{bmatrix} w_0 & & 0 \\ & \ddots & \\ 0 & & w_{a+b} \end{bmatrix} \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \alpha_1 & \cdots & \alpha_1^{a+b} \\ \vdots & & & \\ 1 & \alpha_{a+b} & \cdots & \alpha_{a+b}^{a+b} \end{bmatrix}$$

or

$$W = \begin{bmatrix} w_0 & & 0 \\ & \ddots & \\ 0 & & w_{a+b} \end{bmatrix} \begin{bmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{a+b} \\ 1 & \alpha_1 & \cdots & \alpha_1^{a+b} \\ \vdots & & & \\ 1 & \alpha_{a+b} & \cdots & \alpha_{a+b}^{a+b} \end{bmatrix}.$$

REMARK. We see that the Vandermonde matrix enters into every bilinear minimal algorithm α.

We will now prepare ourselves to prove the following result. Let $P = z^n +$

$\sum_{i=0}^{n-1} g_i z^i$, where $P = \bar{P}^l$ and $\bar{P}$ is irreducible over G. If

$$R = \sum_{i=0}^{n-1} x_i z^i, \qquad S = \sum_{i=0}^{n-1} y_i z^i,$$

then the system of bilinear forms T_p has m/d number $2n - 1$.

Let C_P be the companion matrix of P acting on the column vector space V^*.

$$C_P = \begin{bmatrix} 0 & \cdots & -g_0 \\ 1 & & \vdots \\ & \ddots & \vdots \\ 0 & & 1 - g_{n-1} \end{bmatrix}.$$

The minimal polynomial of C_P is P itself. This means that $P(C_P)$ is the zero matrix and any other polynomial with this property is divisible by P. Now let $v \in V$, $v \neq 0$. Consider the set $\mathfrak{I}$ of polynomial Q such that

$$vQ(C_P) = 0.$$

Clearly $\mathfrak{I}$ is an ideal, $\mathfrak{I} \supset P$. Since C_P is nonsingular $z \notin \mathfrak{I}$ and $\mathfrak{I} \neq G[z]$. Let $(\bar{P})$ be the ideal generated by $\bar{P}$. If B and P are relatively prime then the ideal generated by B and P is $G[x]$. Since $\mathfrak{I} \neq G[z]$, $\mathfrak{I} \subset (\bar{P})$. Hence if Q is not divisible by $\bar{P}$, $vQ(C_P) \neq 0$ any $v \in V^*$ and so $Q(C_P)$ is nonsingular.

LEMMA III.4.2. *Let* $V_P = \{v \in V |$ *there exists a polynomial* Q *of degree* $< n$ *and* $vQ(C_P) = 0\}$. *Then* $\dim V_P < n$.

PROOF. Let $W = \{v \in V | v\bar{P}^{l-1}(C_P) = 0\}$. Claim $W = V_P$. Clearly $W \subset V_P$. If $vQ(C_P) = 0$, then by the above discussion

$$Q = \bar{P}^r Q', \qquad Q' \text{ relatively prime to } \bar{P},$$

where $l > r > 0$ and $Q'(C_P)$ is nonsingular. But then $v\bar{P}^r(C_P) = 0$ which implies $v\bar{P}^{l-1}(C_P) = 0$.

LEMMA III.4.3. *Let* C_P *be the companion matrix to* $P(z)$. *If*

$$t = \begin{bmatrix} t_0 \\ \vdots \\ t_{n-1} \end{bmatrix}$$

then the coefficients of $z \sum_{i=0}^{n-1} t_i z^i \bmod P(z)$ *are* $C_P(t)$.

This is essentially the definition of C_P.

LEMMA III.4.4. *Let* $R(z) = \sum_{i=0}^{n-1} x_i z^i$ *and* $S(z) = \sum_{j=0}^{n-1} y_j z^j$ *and let* $\tilde{T}_P$ *be the system of bilinear forms that are the coefficients of* $R(z) \cdot S(z) \bmod P(z)$. *Let* $T_P = A(x)y$ *as in* §*III*.3. *Then*

$$A(x) = \left(X, C_P X, \ldots, C_P^{n-1} X\right)$$

where

$$X = \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix},$$

C_P is the companion matrix to P, and C_P^α is the α power of the matrix C_P.

PROOF. We have, since $R(z) = \sum x_i z^i$ and $S(z) = \sum y_j z^j$, that

$$S(z) \cdot R(z) = y_0 \sum x_i z^i + y_1 z \sum x_i z^i + \cdots + y_{n-1} z^{n-1} \sum x_i z^i.$$

By Lemma III.4.3 we have

$$S(z) \cdot R(z) = y_0 X + y_1 C_P X + \cdots + y_{n-1} C_P^{n-1} X.$$

The coefficients of z^k in $S(z) \cdot R(z)$ is $\sum y_k \xi_k$, where ξ_k is the k entry in the column vector $C_P^k X$, $k = 0, \ldots, n-1$.

We are now in a position to prove the following theorem.

THEOREM III.4.5. *Let $R(z)$ and $S(z)$ be polynomials of degree $n-1$. Let $P = \bar{P}^l$ where $\bar{P}$ is irreducible over G and let* deg $P = n$. *The minimum number of multiplications needed to compute $\tilde{T}_P$ is $2n-1$, where $\tilde{T}_P$ is the system of bilinear forms that are the coefficients of $R(z) \cdot S(z)$* mod $P(z)$.

PROOF. Let r be the minimum number of multiplications needed to compute $\tilde{T}_P$. By Theorem III.3.3, we have

$$A(x)y = Um$$

where U is an $n \times r$ matrix over G and

$$m = \begin{bmatrix} L_1 \cdot L_1^1 \\ \vdots \\ L_l \cdot L_l^1 \end{bmatrix}.$$

Let V be an n-dimensional G-vector space, let V^* be its dual space and let $w \in V^*$. Then $wA(x) \not\equiv 0$ because its first coefficient is

$$\sum w_i x_i.$$

Hence wUm is not zero and so wU is not zero. Since w was arbitrary, this shows that the rank of U is n. By reordering columns, if necessary, we may assume that we have a nonsingular $n \times n$ matrix W since

$$WU = (I | U')$$

where I is the $n \times n$ identity matrix.

Let V_P be as in Lemma III.4.2. Then, because W is nonsingular, there exists a row of W, say the first, denoted by w, which is not in V_P. Then $wA(x)y$ is a bilinear form and

$$wA(x)y = (1\,0 \ldots 0\, u_1^1 \ldots u_{r-n}^1)m.$$

Thus the bilinear form on the left, above, can be computed using $r - n + 1$ multiplications. We now claim that the n columns of $wA(x)$ are independent.

This would imply our theorem for by Theorem III.3.2 $r - n + 1 \geqslant n$ or $r \geqslant 2n - 1$.

To show that the columns of $wA(x)$ are independent, assume that

$$0 = \sum_{i=0}^{n-1} wC_P^i X \cdot \alpha_i = w\left(\sum_{i=0}^{n-1} \alpha_i C_P^i\right) X.$$

Since the elements of X are indeterminants over G

$$w\left(\sum_{i=0}^{n-1} \alpha_i C_P^i\right) = 0.$$

But $w \notin V_P$ and the above contradicts Lemma III.4.2.

COROLLARY III.4.6. *Let $P = \bar{P}_1^{l_1}, \ldots, \bar{P}_k^{l_k}$ where $\bar{P}_i$ are distinct irreducible polynomials over G. Then $\tilde{T}_P$ can be computed in $2n - k$ multiplications and no divisions.*

This follows easily from the discussion at the beginning of this section.

Before coming to the final result to be proven in this paper, we would like to remind the reader of the relation between the Chinese Remainder Theorem and the rational form theorem for a matrix.

Again let $P = P_1, \ldots, P_k$ where $P_i = \bar{P}_i^{l_i}$ and the $\bar{P}_i$ are distinct irreducible polynomials in $G[z]$. Let C_P be the companion matrix to P acting on V^*. By the Chinese Remainder Theorem

$$G[z]/P = \sum \bigoplus_{i=1}^{k} G(z)/P_i$$

where equality denotes isomorphic. The rational canonical form theorem says that there exists subspaces V^* of V^* such that

(a) $V^* = \Sigma \oplus V_i^*$,

(b) $C_P(V_i^*) = V_i^*$,

(c) $C_P | V_i^* = C_{P_i}$,

where C_{P_i} is the companion matrix to P_i.

THEOREM III.4.7. *Let $P = P_1, \ldots, P_k$ where $P_i = \bar{P}^{l_i}$ and $\bar{P}_i$ are distinct irreducible polynomials in $G[z]$. Further let the degree of $P_i = n_i$ and $n = \Sigma_{i=1}^k n_i$. Let $R(z) = \Sigma_{i=0}^{n-1} x_i z^i$ and $S(z) = \Sigma_{i=j}^{n-1} y_i z^i$. If $T = R \cdot S$ then $\tilde{T}_P$ cannot be computed with less than $2n - k$ essential multiplications and no divisions.*

We have already seen in Corollary III.4.6 that the m/d number is bounded above by $2n - k$. In the proof of the theorem we will need the following purely technical lemma. We will state and prove this lemma below, but the reader may prefer to skip directly to the proof of Theorem III.4.7 and return to the lemma later.

LEMMA III.4.8. *Let $P = P_1, \ldots, P_k$ be as in Theorem* III.4.7 *and let V_{P_i}, $i = 1, \ldots, k$, be as in Lemma* III.4.2. *Let W be a nonsingular $n \times n$ matrix and let W^1 be the first n_1 columns of W, let W^2 be the next n_2 columns of W, etc. Then each W^i has a row $w^i(j(i))$, j a function of i, such that $w^i(j(i)) \notin V_{P_i}$. Further, there exists a $1 \times n$ matrix β with nonzero entries only at the $j(i)$,*

$i = 1, \ldots, k$, coordinates such that

$$\beta W = (\gamma_1, \ldots, \gamma_k)$$

where γ_j is a $1 \times n_j$ matrix $j = 1, \ldots, k$ and $\gamma_j \notin V_{P_j}$.

PROOF. Since W is nonsingular, each matrix W^j, $j = 1, \ldots, k$, has rank n_j. By Lemma III.4.2 there must be a row $w^i(j(i)) \notin V_{P_i}$. Consider the set U of $1 \times n$ vectors with nonzero entries only at $j(i)$, $i = 1, \ldots, k$. U is a k-dimensional vector space.

In W^i consider the row vectors $w^i(j(i))$, $j(i) = 1, \ldots, k$, and form $\Sigma_{i=1}^k \beta_{j(i)} w^i(j(i))$. The set $U(i)$ of β such that

$$\sum_{i=1}^{k} \beta_{j(i)} w^i(j(i)) \notin V_{P_i}$$

is the complement of a proper linear subspace in U. Hence $\bigcap_{i=1}^k U(i)$ is not empty. Any point in $\bigcap U(i)$ satisfies the conclusion of Lemma III.4.8.

PROOF OF THEOREM. By our discussion about the Chinese Remainder Theorem and Rational Canonical Form Theorem, we see that we may assume (without use of essential m/d) that $\tilde{T}_{P_i}$ corresponds to $A_i(x^i)y^i$ and $\tilde{T}$ to $A(x)y$ where

$$A(x) = \begin{bmatrix} A_1(x^1) & & 0 \\ & \ddots & \\ 0 & & A_k(x^k) \end{bmatrix}, \qquad y = \begin{bmatrix} y^1 \\ \vdots \\ y^k \end{bmatrix}.$$

Let $A(x)y = Um$, by Theorem III.3.3. Then U is an $n \times t$ matrix over G where t is the number of essential multiplications. Since all rows of $A(x)$ are linearly independent, the rank of U is n. Therefore, there exists an $n \times n$ nonsingular matrix W such that

$$WU = (I|U'), \quad \text{where } I \text{ is the } n \times n \text{ identity matrix.}$$

Applying Lemma III.4.8 to W and letting β be as in Lemma III.4.8, we consider the bilinear form

$$\beta WA(x)y = \beta(I|U')m.$$

We claim that at least n multiplications are needed to compute this bilinear form. We prove this by showing that the column rank of the left side above is n. Let $\gamma = \beta W$ and consider every nontrivial linear combination of the column of $\beta WA(x) = \gamma A(x)$. Substituting $C_{P_i}^j x^i$ for the j element of $A_i(x^i)$ we obtain this linear combination as

$$\sum \gamma_i \Big(\sum \alpha_{ij} C_{P_i}^j \Big) x^i.$$

This vanishes only if $\gamma_i \Sigma\, \alpha_{ij} C_{P_i}^j = 0$ for $i = 1, \ldots, k$. But $\gamma_i \notin V_{P_i}$ and so $\alpha_{ij} = 0$ for all i and j. Thus by Theorem III.3.2 it requires at least n essential multiplications to compute $\gamma A(x)y$. But $\beta(I|U')$ has at most $k + t - n$ nonzero coefficients. Hence

$$k + t - n \geqslant n \quad \text{or} \quad t \geqslant 2n - k$$

and our theorem is proven.

5. Algorithms for computing the finite Fourier transform. In §III.2 we presented S. Winograd's algorithm for computing $F(7)$. We are now in a position to outline the steps for creating that algorithm as discussed in [21]. However, before proceeding to this, let us at least mention a fundamental result proven in [24].

THEOREM III.5.1. *The m/d number for computing the finite Fourier transform on a prime number p is $2p - 3 - \xi(p-1)$, where $\xi(n)$ is the number of d such that $d|n$.*

We cannot discuss the proof of this theorem in this paper as it is too long. It does show that the algorithm of §III.2 which uses 8 multiplications achieves the minimum number of multiplications for computing $F(7)$.

We will now outline the steps that are followed in creating the algorithm for $F(7)$.

Step 1. Consider the cyclic group $(\mathbf{Z}/7)^x$ of order 6 and show that $e^{2\pi i 3/7} = \omega^3$ is a generator of $(\mathbf{Z}/7)^x$ and that $\omega^1, \omega^3, \omega^2, \omega^6, \omega^4, \omega^5$, correspond to $1, (\omega^3), (\omega^3)^2, (\omega^3)^3, (\omega^3)^4, (\omega^3)^5$.

Step 2. This enables us to show that the Fourier transform can be computed from the coefficients of the polynomial

$$R(z)\cdot S(z) \mod z^6 - 1 = (\omega^1 + \omega^3 z + \omega^2 z^2 + \omega^6 z^3 + \omega^4 z^4 + \omega^5 z^5) \cdot (a_1 + a_5 z + a_4 z^2 + a_6 z^3 + a_2 z^4 + a_3 z^5) \mod z^6 - 1$$

in the notation of §III.2.

Step 3. We verify that

$$x^6 - 1 = (x-1)(x+1)(x^2+x+1)(x^2-x+1) = P_1, P_2, P_3, P_4.$$

Step 4. If

$$Q_i = \delta_{ij} \mod P_j, \qquad i,j = 1, 2, 3, 4,$$

as in §III.4, we compute and find

$$Q_1 = \frac{1}{6}(x+1)(x^2+x+1)(x^2-x+1),$$

$$Q_2 = -\frac{1}{6}(x-1)(x^2+x+1)(x^2-x+1),$$

$$Q_3 = -\left(\frac{x}{6}+\frac{1}{3}\right)(x^2-1)(x^2-x+1),$$

$$Q_4 = -\left(\frac{x}{6}-\frac{1}{3}\right)(x^2-1)(x^2+x+1).$$

Step 5. Compute

$$\begin{matrix} R \bmod P_i, \\ S \bmod P_i, \end{matrix} \qquad i = 1, 2, 3, 4,$$

and so compute

$$R\cdot S \bmod z^6 - 1 = \sum Q_i (R \bmod P_i)(S \bmod P_i) \mod z^6 - 1.$$

Step 6. Apply the Chinese Remainder Theorem twice more to compute

$$R_3S_3 \bmod P_3 \quad \text{and} \quad R_4S_4 \bmod P_4$$

where

$$R_3 = R \bmod P_3 \quad \text{and} \quad R_4 = R \bmod P_4$$
$$S_3 = S \bmod P_3 \quad \text{and} \quad S_4 = S \bmod P_4.$$

Step 7. Combine all the previous computations into a bilinear algorithm.

REFERENCES

1. L. Auslander, *Lecture notes on nil-theta functions*, CBMS Regional Conf. Ser. in Math. no. 34, Amer. Math. Soc., Providence, R. I., 1977.

2. L. Auslander and J. Brezin, *Translation invariant subspaces in L^2 of a compact nilmanifold.* I, Invent. Math. **20** (1973), 1–14.

3. L. Auslander and R. Tolimieri (assisted by H. E. Rauch), *Abelian harmonic analysis, theta functions and function algebras on a nilmanifold*, Lecture Notes in Math., vol. 436, Springer-Verlag, Berlin and New York, 1975.

4. _____, *Algebraic structures for $\bigoplus \Sigma_{n>1} L^2(Z/n)$ compatible with the finite Fourier transform*, Trans. Amer. Math. Soc. **244** (1978), 263–272.

5. R. Bellman, *A brief introduction to theta functions*, Holt, Rinehart and Winston, New York, 1961.

6. Z. I. Borevich and I. R. Shafarevich, *Number theory*, Academic Press, New York, 1966.

7. J. Brezin, *Harmonic analysis on nilmanifolds*, Trans. Amer. Math. Soc. **150** (1970), 611–618.

8. _____, *Harmonic analysis on compact solvmanifolds*, Lecture Notes in Math., vol. 602, Springer-Verlag, Berlin and New York, 1977.

9. J. W. Cooley, P. A. W. Lewis, and P. P. Welch, *Historical notes on the fast Fourier transform*, Proc. IEEE **55** (1967), 1675–1677.

10. J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, Math. Comput. **19** (1965), 297–301.

11. I. J. Good, *Analogues of Poisson's summation formula*, Amer. Math. Monthly **69** (1962), 259–266.

12. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Clarendon Press, Oxford, 1938.

13. K. Ireland and M. I. Rosen, *Elements of number theory*, Bogden and Quigley, New York, 1972.

14. S. Lang, *Algebraic number theory*, Addison-Wesley, Reading, Mass., 1970.

15. J. H. McClellan, *Comments on "eigenvector and eigenvalue decomposition of the discrete Fourier transform"*,IEEE Trans. Audio and Electroacoust. (1972), 65.

16. J. H. McClellan and T. W. Parks, *Eigenvalue and eigenvector decomposition of the discrete Fourier transform*, IEEE Trans. Audio and Electroacoust. March (1972), 66–74.

17. H. Rademacher, *Lectures on elementary number theory*, Blaisdell, Boston, Mass., 1964.

18. R. Tolimieri, *The multiplicity problem for 4-dimensional solvmanifolds*, Bull. Amer. Math. Soc. **83** (1977), 365–366.

19. A. Weil, *Sur certaines groupes d'operateurs unitaires*, Acta Math. **111** (1964), 143–211.

20. S. Winograd, *On computing the discrete Fourier transform*, Proc. Nat. Acad. Sci. U.S.A. **73** (1976), 1005–1006.

21. _____, *On computing the discrete Fourier transform*, I.B.M. Research Report, 1976.

22. _____, *On the number of multiplications necessary to compute certain functions*, Comm. Pure Appl. Math. **23** (1970), 165–179.

23. _____, *Some bilinear forms whose multiplicative complexity depends on the field of constants*, Math. Systems Theory **10** (1977), 169–180.

24. _____, *On the multiplicative complexity of the discrete Fourier transform*, Advances in Math. (to appear).

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL AND UNIVERSITY CENTER (CUNY), NEW YORK, NEW YORK 10036

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CONNECTICUT, STORRS, CONNECTICUT 06268