[redacted]

1973

On the Arithmetic Complexity of a Class of Arithmetic Computations

David Paul Dobkin

Research Report #23

This work is essentially the same as the author's doctoral dissertation, prepared for the Division of Engineering and Applied Physics at Harvard University, 1973.

October 1973

[redacted]

PREFACE

I am extremely grateful to Professor Roger W. Brockett for his continual guidance and encouragement during the course of this research. His willingness to take me on as a student and his eagerness to learn about arithmetic complexity during the course of this research contributed greatly to the quality of this thesis. As a teacher, colleague and friend he has had a profound affect on my graduate life.

I would also like to thank my friends and fellow graduate students for many useful diversions and some helpful technical discussions during the last three years. The helpful comments of Professors Ronald V. Book and Donald J. Rose on the rough draft have led to a much clearer final form of this manuscript. I would also like to thank Ms. Maureen Stanton for her excellent typing of the main body of this dissertation and Ms. Frances Gedzium for helping with the final details.

Last, but not least, I am most grateful to my wife Kathy who has contributed to this dissertation in other ways. Her continual support, willingness to help me understand myself and her general attitude have contributed to my intellectual growth as well as to my personal growth.

Portions of this research were performed under support given to Harvard by the U.S. Office of Naval Research under the Joint Services Electronics Program by Contract N00014-67-A-0298-0006 and the National Science Foundation by Grant GJ-30076.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYNOPSIS

Numerous studies have been made concerning the complexity of arithmetic operations. However, no general theory has emerged for studying classes of operations and very few general methods exist for determining the complexity of a given operation. Such a general theory would be useful in unifying the theory of arithmetic complexity by identifying the various structures determining the complexity of an operation. If the many ad hoc methods that have been previously used could be formalized within this theory, then general methods for obtaining bounds in the theory of arithmetic complexity would result.

In this thesis, a general model is proposed for studying bilinear multiplication operations in order to provide a common framework for discussing problems regarding a wide class of arithmetic operations. Analysis and synthesis methods are given within the framework to yield methods of obtaining upper and lower bounds on the complexity of operations in this class. Extensions of the model and analysis methods to n-linear operations are also studied.

With this general model, a number of complexity problems are reduced to a problem in linear algebra relating to the expansion of a given set of matrices as linear combinations of rank one matrices. A systematic attack on this problem is made here and some general results are derived which unify and extend numerous known results. Some important problems in arithmetic complexity are reduced to some basic questions in tensor analysis through this model. On the basis of this reduction, synthesis methods are related to tensor ranking methods in an extension of a few of the fundamental concepts of matrix algebra.

A number of new results are given here to illustrate the strength of this approach. Among these is a new lower bound on the number of multiplications required for n by n matrix multiplication of $3n^2-3n+1$ which is independent of the subset of the complex field with regard to which multiplication is regarded as free. An even sharper bound is obtained if this set is restricted to the integers. The results of studies of polynomial multiplication with multiplication by integers regarded as free are also included. These studies, in which connections are established with research in algebraic coding theory, represent an improvement over previous results. In the study of n-linear operations, connections between determinant and permanent computations on a matrix are studied and an interesting open question is proposed. Suggestions for further research are included and consider the possibility of establishing connections between the results presented and successful theories in closely related fields of study.

Portions of the research reported here represent joint work with R. Brockett and have previously appeared as [3].

CHAPTER I

INTRODUCTION

Arithmetic complexity is concerned with the difficulty of evaluating a function at a set of points. The given data in an arithmetic complexity problem is a description of the function to be evaluated and of the set of points at which the function is to be evaluated. The goal is to find the algorithm, within a given class of algorithms, that can be used to evaluate the function most efficiently, subject to a given cost criterion. Typical of the functions to be evaluated in arithmetic complexity are basic multiplication operations corresponding to computations of products of matrices, polynomials and integers, of determinants and permanents of matrices and of the values of a function and its derivatives at a point or set of points [1,15] and cost criteria generally consist of weighted sums of elementary operaion counts. The relative ease with which arithmetic complexity problems can be stated leads to an anomalous situation, since some of the most easily stated problems remain unsolved. For example, although it is possible to prove that every positive integer is expressible as the sum of the squares of four integers, the best known results about the minimum number of multiplications necessary to multiply two 3 x 3 matrices put this number between 19 and 24.

The operations studied in a theory of arithmetic complexity may be represented as mappings between sets. Then the structures of these mappings identify classes of operations which can be studied together. In this dissertation, the class of n-linear multiplication operations is studied. This class includes the operations of multiplying sets of matrices, polynomials, quaternions and complex numbers as well as determinant and permanent calculations on n by n matrices. If the elements of the inputs and outputs for a given n-linear operation belong to the ring $\mathscr{R}$, then the input set consists of n-tuples, the ith of whose elements is a p_i-triple over $\mathscr{R}$, and the output set of p_o-tuples over $\mathscr{R}$. In this case, the input-output map defining the operation is a linear function of each of the entries of the input n-tuples. An important subclass of the class of n-linear operations is the class of bilinear operations which are studied in great detail here.

The purpose of this dissertation is to investigate the mathematical structures associated with bilinear operations in order to determine the basic mathematical questions involved. The formalism developed is used to arrive at new results on the number of operations required to do specific types of problems. These results follow from the unification of ad hoc methods which have been used previously to obtain complexity bounds on these operations. This unification results in general methods of determining the complexity of evaluating any bilinear multiplication operation with respect to any cost criterion. Furthermore, we are able to relate many of the questions asked in this framework to problems in the structure of tensors of arbitrary order, distance properties of vectors in a finite dimensional vector space and various other algebraic and combinatorial problems which remain unsolved.

The basic problem involved in studying bilinear operations is to find methods of evaluating the set $\{<x,G_i\, y> \mid i=1,\ldots,m\}$ of bilinear forms at arbitrary pairs of points (x,y) by a method requiring a minimal number of operations. In this formulation x(resp. y) is a p-tuple (resp. q-tuple) over a given ring $\mathscr{R}$, $\{G_i\}$ is a set of p by q matrices with entries in the real field, and $\langle\cdot,\cdot\rangle$ represents the standard inner product. General approaches to this problem [6,22,24,26] have been previously studied as have applications of these and other approaches to determining the complexity of bilinear operations such as matrix multiplication [5,6,9,11,12, 14,21,25,27], polynomial multiplication [2,5,6,15,20] and integer multiplication [7,15,20].

The class of methods considered here for evaluating sets of bilinear forms is similar to those considered in previous approaches. Straight-line algorithms (i.e. algorithms with no branching) consisting of three stages are considered for this evaluation. The first stage in the evaluation is to compute the sets $\{\langle c_j,x\rangle \ , \ j=1,\ldots,d\}$ and $\{\langle b_j,y\rangle \ , \ j=1,\ldots,d\}$ of linear forms over the complex field. The products $p_j = \langle c_j,x\rangle \langle b_j,y\rangle$ are computed in the second stage and in the final stage, linear combinations of the form $\sum_{j=1}^{d} a_{ji}\, p_j = \langle x, G_i\, y\rangle$ are computed. Algorithms of this form were studied by Fiduccia [6], Gastinel [9] and Hopcroft and Musinski [12]. Strassen [22] showed that if we denote the jk element of G_i by h_{ijk} , then

$$h_{ijk} = \sum_{\alpha=1}^{d} a_{\alpha i}\, (d_\alpha)_j\, (b_\alpha)_k$$

where $(c_\alpha)_j$ (resp. $(b_\alpha)_k$) represents the jth (resp. kth) component of the vector c_α (resp. b_α).

An important contribution of this dissertation is the introduction of a total problem formulation extending previous approaches and an analysis of the mathematical issues involved in solving this problem. As stated above, a bilinear multiplication problem gives rise to a set of matrices $\{G_i\}$ and to a set of structural constants (h_{ijk}). We find it convenient to represent these quantities by introducing the set of indeterminates $s_1,\ldots,s_m$ and to define for a bilinear multiplication problem, a characteristic function, $G(s) = \sum_{i=1}^{m} s_i G_i$ and a defining function, $H(s,x,y) = \langle x,G(s)y\rangle$. An algorithm of the form considered above for evaluating $\{\langle x,G_iy\rangle\}$ corresponds to factorizations of the form $H(s,x,y) = \sum_{i=1}^{d} \langle a_i,s\rangle \langle c_i,x\rangle \langle b_i,y\rangle$ and $G(s) = CA(s)B$ where the ith column (resp. row) of the matrix C(resp. B) is c_i(resp. b_i'), a_i represents a d-vector with jth element a_{ij} and $A(s)$ is a diagonal matrix with ii entry $\langle a_i,s\rangle$. A triple of matrices (A,B,C) will be called a realization of $G(s)$ if their rows and columns form a factorizations of $G(s)$ and $H(s,x,y)$ as above. The dimension of a realization is defined as the number of linear forms in the inputs computed, d in the exposition above. The study of realizations and the introduction of indeterminate variables to yield a symmetric problem formulation are among the important new contributions of this formulation.

Within this framework, it is possible to study arithmetic complexity questions for bilinear multiplication problems in a uniform manner. The basic questions proposed and studied here involve the correspondence between realizations of characteristic functions and the complexity of algorithms for evaluating bilinear multiplication operations. We define the complexity of an algorithm for evaluating a bilinear operation as the number of multiplications required to evaluate the operation for worst-case inputs. Previous models have used this criterion [5,6,9,12,17,24,26] or a criterion counting the number of additions required for evaluation [13,14]. Since approaches to the multiplicative and additive complexity of bilinear operations are similar, it is reasonable to assume that analysis results obtained using either of these criteria are applicable to the other and also to criteria involving weighted sums of multiplication and addition operations. Thus, although complexity as used here directly implies multiplicative complexity, extension to broader cost criteria is possible.

As presented here, algorithms for bilinear operations consists of three stages. The first stage of a bilinear algorithm represents the preconditioning stage, the second the computing stage and the third the postconditioning stage. The measure of complexity introduced allows for varying amounts of preconditioning and postconditioning to be performed and not counted. This is achieved by defining a K-realization of $G(s)$ as a triple (A,B,C) which forms a realization of $G(s)$ such that all elements of A , B and C belong to K , a subset of $\mathbb{C}$. The degree of the characteristic function $G(s)$ over K , represented by $\delta_K(G(s))$ is the dimension of the K-realization of $G(s)$ of minimum dimension. This realization gives rise to an algorithm for computing the bilinear operation corresponding to $G(s)$ which is of optimal efficiency over K . The function $\delta_K(G(s))$ evaluated over subjects K of $\mathbb{C}$ gives a total characterization of the complexity of the bilinear operation giving rise to $G(s)$. An important contribution of this dissertation is the allowance for variations in K in computing $\delta_K(G(s))$. Previous formulations were restricted to the cases $K = \mathbb{Z}$ [11,12] or $K = \mathbb{R}$ [6,23].

Perhaps the most important tests of a problem formulation are its effectiveness in modelling a real situation and the simplification of solution which it generates. The model we give is fully defined and its effectiveness defended in Chapter II. Chapter III is devoted to studying methods of achieving the second goal, by generating analysis and synthesis methods for evaluating $\delta_K(G(s))$ for arbitrary K and G(s). In Chapter II, we discuss methods of generating classes of realizations from a given one and methods of comparing the complexities of seemingly unrelating operations are presented, the differences between cost functions (i.e. measures of complexity) are also discussed with regard to practical issues involved with preconditiong. The analysis methods presented in Chapter III are used to evaluate $\delta_K(G(s))$ for arbitrary K by examining the degrees of subcharacteristic functions of G(s) and studying the form of G(s), Methods of finding lower bounds on $\delta_K(G(s))$ by evaluating the degrees of operations which are easier than G(s) are also studied as well as methods of reducing the evaluation of $\delta_{\mathbb{Z}}(G(s))$ to a problem in coding theory. A synthesis procedure for bilinear operations is also studied.

Chapter IV is devoted to understanding extensions of the results obtained for bilinear multiplication problems to an n-linear structure. The obvious extension of the bilinear model is presented, its strengths and weaknesses are evaluated and alternatives are mentioned. Contact is established with Strassen's [23] lower bound of O(n log n) operations for the evaluation of the symmetric functions in n indeterminates. Of particular interest in this setting are the problems of computing the determinant and permanent of an $n \times n$ matrix as functions of its rows or columns. These problems are studied in the context of each of the models proposed for evaluating n-linear operations.

In the final chapter, the results presented within the dissertation are summarized. Various suggestions for further research are given in addition to the open problems and conjectures presented throughout the body of the thesis.

CHAPTER II

THE OPTIMAL EVALUATION OF A SET OF BILINEAR FORMS

1. Bilinear Multiplication Problems.

The major focus of this chapter is on the evaluation of a number of bilinear forms, and in particular, on minimizing the number of multiplications required for such an evaluation. If $\underline{x}$(resp. $\underline{y}$) is a p-vector (resp. q-vector) over an arbitrary field and G_i is a matrix over this field, i=1,...n, then we wish to evaluate the expressions

$$<x,G_i y> = \sum_{k=1}^{q} \sum_{j=1}^{p} h_{ijk} x_j y_k; \quad i=1,2,\ldots m \qquad (1)$$

where $<\cdot,\cdot>$ represents the standard inner product. The approach taken is to adopt a certain general model for the computation and to discuss, within this context, upper and lower bounds on the number of multiplications. These sets of bilinear forms arise in considering bilinear multiplication problems such as matrix multiplication, polynomial multiplication, etc.

The multiplication problem is defined completely by the constants (h_{ijk}) which we hereafter refer to as the structural constants of the multiplication problem. Since we deal with bilinear forms rather than quadratic forms there are no natural symmetries satisfied by the numbers h_{ijk}. Therefore, any such array is potentially interesting set of structural constants. To avoid trivialities we assume that no nontrivial linear combination of the type

$$\Sigma\alpha_k h_{ijk}$$

or

$$\Sigma\beta_j h_{ijk}$$

or

$$\Sigma\gamma_i h_{ijk}$$

vanishes.Such structural constants will be said to define *nondegenerate multiplication problems*. As will become clear, the general problem can be reduced to a nondegenerate one in an obvious way.

It is also convenient to characterize the given problem by the G_i or more briefly by a degree one matrix polynomial in m indeterminants.

$$G(s) \overset{def}{=} \sum_{i=1}^{m} s_i G_i \qquad (2)$$

where the jkth element of G_i is just h_{ijk}. We refer to $G(s)$ as the *characteristic function* of the problem. Finally the scalar

$$H(s,x,y) = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{q} h_{ijk} s_i x_j y_k = \langle x, G(s)y\rangle \qquad (3)$$

is called the *defining function*. The integer m plays a role analogous to p = dim x and q = dim y. We call m the *index* of the problem.

If the G_i can be expressed as a linear combination of rank one matrices

$$G_i = \sum_{\alpha=1}^{n} a_{i\alpha} D_\alpha$$

then the set of bilinear forms (1) can be evaluated with n multiplications. That is, since each of the D_α can be expressed as $D_\alpha = c_\alpha b_\alpha'$ where c_α is a column vector and b_α' (the transpose of b_α) is a row vector

$$\begin{aligned}\langle x, G_i y\rangle &= \sum_{\alpha=1}^{n} a_{i\alpha} \langle x, c_\alpha b_\alpha' y\rangle \\ &= \sum_{\alpha=1}^{n} a_{i\alpha} \langle x, c_\alpha\rangle b_\alpha' y\end{aligned}$$

$$\overset{-}{x} \quad \overset{|}{c_\alpha} \quad \overset{-}{b_\alpha} \quad \overset{|}{y}$$

$$= \sum_{\alpha=1}^{n} a_{i\alpha}\langle c_\alpha,x\rangle\langle b_\alpha,y\rangle \qquad (4)$$

If the $a_{i\alpha}$ and the elements of c_α and b_α are sufficiently simple so as to make multiplication by them negligible i.e. if these multiplications can be regarded as free, then only the products $\langle c_\alpha,x\rangle$ times $\langle b_\alpha,y\rangle$ need by counted in assessing the complexity of the computation. This development is similar to the development of Strassen [22], pp. 9-10], Fiduccia [6, pp. 36-7], Gastinel [9, p. 222] and Hopcroft-Musinski [12, p. 74]. The major new contribution of this development is the introduction of the indeterminates s_i and the consideration of the index (= dim s) of the problem in a manner analogous to p = dim x and q = dim y.

We let K be a subset of the real numbers such that multiplication of any number by an element of K is not to be counted in tallying the number of effective multiplications required. In order for this to be a self-consistent measure of complexity in a setup where addition is free and iteration is permitted K would have to be closed under addition and multiplication. However for most of the mathematical questions which arise here no ambiguity results if we do not require K to be a subring of $\mathbb{R}$ or $\mathbb{C}$ so we do not make this a general hypothesis. In any case, for most of our results the exact choice of K is irrelevant.

In decomposing the G_i as a sum of rank one matrices as indicated above, it is necessary that the vectors b_i and c_i have their elements in K. This leads to the definition of the <u>K-degree</u> of (h_{ijk}), or G(s), as the minimum number n such that there exist vectors b_α and c_α with elements in K, and a matrix $(a_{i\alpha})$ with elements in K satisfying

$$G_i = \sum_{\alpha=1}^{n} a_{i\alpha} c_\alpha b'_\alpha \; ; \quad i=1,2,\ldots,m \tag{5}$$

The K-degree is simply the minimum number n such that (h_{ijk}) is expressible (see Strassen [21], page 10) as

$$(h_{ijk}) = \sum_{\alpha=1}^{n} (a_\alpha)_i (c_\alpha)_j (b_\alpha)_k$$

where $(a_\alpha)_i$ represents the ith component of the vector a_α, etc. We denote the K-degree of a characteristic function by $\delta_K[G(s)]$. In terms of the defining function, the K-degree is the smallest number n such that

$$<x,G(s)y> = \sum_{\alpha=1}^{n} <a_\alpha,s><c_\alpha,x><b_\alpha,y>$$

where the entries in a_α, b_α and c_α are all in K. This form reveals the complete symmetry of the roles played by x, y and s, whereas most of the other formulations tend to mask this very important fact. Thus the introduction of s into the problem -- which we regard as one of the important new aspects of this thesis -- pays off by revealing immediately 6 equivalent problems obtained by permuting x, y, s.

We conclude this section with some examples which serve to illustrate these ideas in several specific cases and to fix some notation required later on in the dissertation. These examples are described in greater detail and further examples are given in the appendix.

<u>Example 1:</u> (Complex number multiplication) Consider the problem of multiplying the complex numbers $(x_1+\sqrt{-1}\ x_2)$ and $(y_1+\sqrt{-1}\ y_2)$. If $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, then we wish to compute $<x,G_1y>$ and $<x,G_2y>$, the real and imaginary parts of the product, where the matrices G_1 and G_2 are given by $G_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $G_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This problem gives rise to the characteristic function

$$I(s) = G_1s_1 + G_2s_2 = \begin{bmatrix} s_1 & s_2 \\ s_2 & -s_1 \end{bmatrix}$$

and defining function

$$H(s,x,y) = <x,I(s)y> = x_1y_1s_1+x_1y_2s_2+x_2y_1s_2-x_2y_2s_1 \ .$$

<u>Example 2:</u> (Two by two matrix multiplication) Consider the problem of 2 by 2 matrix multiplication. Let

$$X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \quad ; \qquad Y = \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix}$$

The problem of computing X Y = Z gives rise to the defining function

$$H(s,x,y) = x_1y_1s_1+x_2y_3s_1+x_1y_2s_2+x_2y_4s_2+x_3y_1s_3+x_4y_3s_3+x_3y_2s_4+x_4y_4s_4$$

This gives a characteristic function

$$G(s) = \begin{bmatrix} s_1 & s_2 & 0 & 0 \\ 0 & 0 & s_1 & s_2 \\ s_3 & s_4 & 0 & 0 \\ 0 & 0 & s_3 & s_4 \end{bmatrix}$$

The fact that $H(s,x,y)$ can also be expressed as $\langle y,\check{G}(x)s\rangle$ with

$$\check{G}(x) = \begin{bmatrix} x_1 & 0 & x_3 & 0 \\ 0 & x_1 & 0 & x_3 \\ x_2 & 0 & x_4 & 0 \\ 0 & x_2 & 0 & x_4 \end{bmatrix}$$

and also $\langle s,\hat{G}(y)x\rangle$ where

$$\hat{G}(y) = \begin{bmatrix} y_1 & y_3 & 0 & 0 \\ y_2 & y_4 & 0 & 0 \\ 0 & 0 & y_1 & y_3 \\ 0 & 0 & y_2 & y_4 \end{bmatrix}$$

gives rise to 3 alternative problems. In this case it happens that these forms are also obtained from each other by permutation of rows and columns.

Example 3: (Polynomial multiplication) Consider multiplying a polynomial of degree μ-1 by a polynomial of degree ν-1. If the polynomials are written as

$$x(t) = \sum_{j=1}^{\mu} x_j t^{j-1}$$

$$y(t) = \sum_{k=1}^{\nu} y_k t^{k-1}$$

then the defining function is

$$H(s,x,y) = \sum_{j=1}^{\mu} \sum_{k=1}^{\nu} s_{j+k-1} x_j y_k$$

and the characteristic function is

$$P_{\mu,\nu}(s) = \begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_\nu \\ s_2 & s_3 & s_4 & \cdots & s_{\nu+1} \\ s_3 & s_4 & s_5 & \cdots & s_{\nu+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_\mu & s_{\mu+1} & s_{\mu+2} & \cdots & s_{\mu+\nu-1} \end{bmatrix}$$

In this case by forming $\check{G}(x)$ and $\hat{G}(y)$, we notice that the same defining function arises in the problem of multiplying a μ by ν Toeplitz matrix by a ν-vector or a $\nu \times \mu$ Toeplitz matrix by a μ-vector. Thus these seemingly different problems are all the same, in a sense to be made precise below.

Example 4: (Matrix Multiplication) Let X and Y be $\mu\times\omega$ and $\omega\times\nu$ matrices respectively corresponding to vectors x and y of length $\mu\omega$ and $\omega\nu$ respectively such that the ij element of X is the $i+(j-1)\omega$ element of x and the ij element of Y is the $(i-1)\omega+j$ element of y. If Z is the $\mu\times\nu$ matrix corresponding to the product of X and Y and z is a $\mu\nu$ vector such that the ij element of Z is the $(i-1)\mu+j$ element of z, then the defining function for $\mu\times\omega$ by $\omega\times\nu$ matrix multiplication is given by

$$H(s,x,y) = \sum_{j=1}^{\mu\nu} s_j z_j = \sum_{\alpha=1}^{\nu} \sum_{\alpha=1}^{\nu} s_{(\alpha-1)\mu+\beta} \sum_{\gamma=1}^{\omega} x_{\alpha+(\gamma-1)\omega} y_{(\gamma-1)\omega+\beta}$$

In order to represent the characteristic function for this problem, we introduce the characteristic function

$$\theta_{\nu\mu}(s) = \begin{bmatrix} s_1 & s_2 & \cdots & s_\nu \\ s_{\nu+1} & s_{\nu+2} & \cdots & s_{2\nu} \\ \vdots & \vdots & & \vdots \\ s_{\nu(\mu-1)+1} & s_{\nu(\mu-1)+2} & \cdots & s_{\mu\nu} \end{bmatrix}$$

and represent the characteristic function for μ×ν by ν×ω matrix multiplication by

$$M_{\mu\nu\omega}(s) = \begin{bmatrix} \theta_{\mu\nu}(s) & & & \\ & \theta_{\mu\nu}(s) & & \\ & & \ddots & \\ & & & \theta_{\mu\nu}(s) \end{bmatrix} \quad ; \omega \text{ blocks}$$

As noted by Hopcroft-Musinski [12], permuting the order of the subscripts on M yields problems of equivalent difficulty. In the special case where $\nu=\mu=\omega$, we write $M_n(s)$ instead of $M_{n,n,n}(s)$.

2. Equivalent Characteristic Functions

In order to classify the difficulty of multiplication problems it is desirable to understand the different ways in which problems of the same difficulty can present themselves. With this goal in mind, we introduce a partial order on the set of characteristic functions and explore its usefulness.

Let us indicate by $K^{n \times m}$ the set n by m matrices whose elements take on values in K. We say that a characteristic function $G_1(s)$ <u>K-dominates</u> a characteristic function $G_2(t)$ if there exist matrices P, Q and R having elements in K such that

$$PG_1(Qt)R = G_2(t) \qquad G_1(s) \underset{K}{\supset} G_2(t)$$

This amounts to being able to encode computations of G_2 into computations of G_1 and in this case we write $G_1(s) \underset{K}{\supset} G_2(t)$. We say that $G_1(s)$ and $G_2(t)$ are <u>K-equivalent</u> if $G_1(s) \underset{K}{\supset} G2(t)$, $G_2(t) \underset{K}{\supset} G_1(s)$. We write this as $G_1(s) \underset{K}{\sim} G_2(t)$. We say that two sets of structural constants (h_{ijk}) and (h^*_{ijk}) are <u>K-equivalent</u> if their associated characteristic functions are <u>K-equivalent</u>.

The following theorem justifies our choice of words.

<u>Theorem 1</u>: If $G_1(s)$ K-dominates $G_2(t)$ then $\delta_K[G_1(s)] > \delta_K[G_2(t)]$ and if $G_1(s)$ is K-equivalent to $G_2(t)$ then $\delta_K[G_1(s)] = \delta_K[G_2(t)]$.

<u>Proof</u>: If $G_1(s) = \sum_{i=1}^{m} G_{1i}s_i$ is of degree δ_K, then there exist rank one matrices D_i and a matrix A with ijth entry α_{ij} such that $G_{1i} = \sum_{j=1}^{\delta_K} \alpha_{ij}D_j$ and

$$G_1(s) = \sum_{i=1}^{m} \sum_{j=1}^{\delta_K} s_i\alpha_{ij}D_j$$

By hypothesis $G_2(t) = PG_1(Qt)R$ so that $G_2(t) = \sum_{\ell=1}^{r} G_{2\ell}t_\ell = \sum_{\ell=1}^{r} (\sum_{i=1}^{m} q_{i\ell}PG_{1i}R)t_\ell$ Thus, a realization of $G_2(t)$ of degree δ_K is given by

$$G_2(t) = \sum_{\ell=1}^{r} \sum_{j=1}^{\delta_K} \sum_{i=1}^{m} q_{i\ell}\alpha_{ij}PD_jRt_\ell$$

which shows that $\delta_K(G_2(t) \leq \delta_K = \delta_K(G_1(t))$. The theorem follows obviously from this realization.

There is a second kind of equivalence which relates to the structural constants (h_{ijk}). Notice that $<x,G(s)y> = <y,G^T(s)x>$. We define $\hat{G}$ and $\check{G}$ by the equations

$$\begin{aligned} H(s,x,y) &= <x,G(s)y> \\ &= <s,\hat{G}(y)x> \\ &= <y,\check{G}(x)s> \end{aligned}$$

Two characteristic functions will be said to be permutation-equivalent if they become identical after an interchange of the indices of their structural constants. If we denote by T the map that takes $G(s)$ into $G^T(s)$ and by $\vee$ the map that takes $G(s)$ into $\check{G}(x)$ then T and $\vee$ generate a group which is isomorphic to the permutation group on 3 letters. We denote this group by Γ and its elements by e (the identity), T, $\vee$, $\vee^2=\Lambda$, $T\vee = \vee^2T=\Lambda T$, and $\vee T=T\vee^2 = T\Lambda$. Notice that nondegeneracy is invariant under the action of Γ.

Theorem 2: If $G_1(s)$ and $G_2(t)$ are permutation equivalent they have the same degree. That is $G(s)$, $\check{G}(x)$, $\hat{G}(y)$, and their respective transposes all have the same degree.

Proof: This is an immediate consequence of the fact that equal defining functions have the same degree.

Special cases of permutation equivalence have been studied by Winograd [26] and Fiduccia [6] (matrix-vector products) and Hopcroft-Musinski [12] (matrix multiplication). We find it desirable to formalize this idea here because to leave it implicit would obscure some of the important features of our development.

We close this section with some basic elementary degree estimates.

Theorem 3: Let G(s) be of index m and suppose G(s) is nondegenerate and p by q. Then

i) $\max(m,p,q) \leq \delta_K[G(s)] \leq \min(mp,mq,pq)$

ii) $\max(\text{rank}\,(G_i,\check{G}_i,\hat{G}_i)) \leq \delta_K[G(s)]$

$$\leq \min\Big(\sum_{i=1}^{m} \text{rank } G_i,\ \sum_{i=1}^{p} \text{rank } \check{G}_i,\ \sum_{i=1}^{q} \text{rank } \hat{G}_i\Big)$$

Proof: To establish the first part we observe that we need at least m terms on the right of equation (5) if there are m linearly independent ones on the left. On the other hand, if the G_i have p rows and q columns then they can be expressed as the sum of min(pm,qm,pq) rank one matrices. The other inequalities follow from permutation equivalence as in Theorem 2.

The second statement follows from a slight modification of this same argument.

3. Symmetries of Characteristic Functions

In the previous section, we defined equivalence for characteristic functions. In this section we discuss characteristic functions which are equivalent to themselves. We will make application of these ideas in section 4 which is devoted to the realization of characteristic functions.

In a case such as M_n (Example 4) where a characteristic function is equivalent to characteristic functions to which it is permutation equivalent, the situation is especially interesting as is outlined below.

To begin with we define the stabilizer, Σ_K, of a characteristic function as follows

$$\Sigma_K(G(s)) \overset{\text{def}}{=} \{(P_1,P_2,P_3) | P_1^{-1},P_2^{-1},P_3^{-1} \text{ exist over } K$$
$$P_1,P_2,P_3 \text{ are matrices over } K \text{ and } P_1G(P_2s)P_3 = G(s)\}$$

When K is a subfield of $\mathbb{C}$ it is easy to verify that the stabilizer admits a group structure with the multiplication operation $(P_1,P_2,P_3)\cdot(Q_1,Q_2,Q_3) = (Q_1P_1,P_2Q_2,P_3Q_3)$ and inverse $(P_1,P_2,P_3)^{-1} = (P_1^{-1},P_2^{-1},P_3^{-1})$. The following theorem describes the invariance of this group under the action of Γ.

Theorem 4: Let K be a subfield of $\mathbb{C}$. Then $\Sigma_K(G(s))$ is invariant under the action of the permutation equivalence group Γ; i.e. the six groups $\Sigma_K(G(s))$, $\Sigma_K(\hat{G}(y))$, $\Sigma_K(\check{G}(x))$, $\Sigma_K(G^T(s))$, $\Sigma_K(\hat{G}^T(y))$ and $\Sigma_K(\check{G}^T(x))$ are all isomorphic.

Proof: Let $T : (P_1,P_2,P_3) \to (P_3^T,P_2,P_1^T)$

$V : (P_1,P_2,P_3) \to (P_3^T,P_1^T,P_2)$

if $\gamma \in \Gamma$ and $P \in \Sigma_K(G(s))$, then $\gamma(P) \in \Sigma_K(\gamma(G(s)))$. It is easily verified that γ defines an isomorphism between $\Sigma_K(G(s))$ and $\Sigma_K(\gamma G(s))$. The latter is $\Sigma_K(\hat{G}(y))$ if $\gamma = V^2$. Similar arguments cover the other cases.

To make this idea more concrete we now compute the stabilizer in two important cases.

Lemma: $A\theta_{n,n}(s) = \theta_{n,n}(Bs)C$ where A, B and C are invertible if and only if $B^T = A^T \otimes C^{-1}$.[1]

Proof: Let B^T be partitioned as

$$B^T = \begin{bmatrix} \Delta_{11} & \Delta_{12} & \cdots & \Delta_{1n} \\ \Delta_{21} & \Delta_{22} & \cdots & \Delta_{2n} \\ \vdots & & & \\ \Delta_{n1} & \Delta_{n2} & \cdots & \Delta_{nn} \end{bmatrix}$$

then $A\theta_{n,n}(s) = \theta_{n,n}(Bs)C$ if and only if

$$\sum_{j=1}^{n} a_{ij} s_{(j-1)n+k} = \sum_{\gamma=1}^{n} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} b_{(i-1)n+\gamma,(\alpha-1)n+\beta}\, c_{\gamma k}\, s_{(\alpha-1)n+\beta} \text{ for all } i,k.$$

Equating coefficients for s_p, $p=1,\ldots n^2$ yields

$$\Delta_{ij}C = a_{ji}I_n \quad \text{or} \quad \Delta_{ij} = a_{ji}C^{-1} \text{ and thus } B^T = A^T \otimes C^{-1}.$$

Corollary: If $A\theta_{n,n}(s) = \theta_{n,n}(Bs)C$ where B is a given invertible $n^2 \times n^2$ matrix, then if $B^T = A^T \otimes C^{-1}$ is a factorization of B^T, all others are of the form $B^T = (kA)^T \otimes (kC)^{-1}$ (for $k \neq 0$).

Proof: Obvious.

Lemma: $AM_n(s) = M_n(Bs)C$ if and only if there exists an invertible matrix P such that $A^T = P \otimes A_{11}^T$, $C^{-1} = P^{-1} \otimes C_{11}^{-1}$ and $A_{11}^T \otimes C_{11}^{-1} = B$.

Proof: Let A and C be partitioned into nxn blocks, then $AM_n(s) = M_n(Bs)C$ if and only if

$$A_{ij}\theta_{n,n}(s) = \theta_{n,n}(Bs)C_{ij}$$

Therefore,

$$B^T = (A_{11}^T \otimes C_{11}^{-1}) = (A_{12}^T \otimes C_{12}^{-1}) = \ldots = (A_{nn}^T \otimes C_{nn}^{-1})$$

[1] $\otimes$ represents the Kronecker product of matrices

and

$$A^T_{ij} = P_{ij}A^T \quad , \quad C^{-1}_{ij} = \frac{1}{P_{ij}} C^{-1}_{11}, \; C_{ij} = P_{ij}C_{11}$$

This yields the conclusions

$$A^T = P \otimes A^T_{11}, \; C = P \otimes C_{11} \text{ and } B = A^T_{11} \otimes C^{-1}_{11} \text{ for an}$$

invertible matrix P.

These results can now be summarized in a statement of the stabilizer $\Sigma_K(M_n(s))$.

<u>Theorem 5</u>: $(A,B,C) \,\varepsilon\, \Sigma_K(M_n(s))$ if and only if there exists an invertible matrix P such that $A = (P^{-1})^T \otimes A^{-1}_{11}$, $C = P \otimes C_{11}$ and $B = A^T_{11} \otimes C^{-1}_{11}$.

<u>Theorem 6</u>: If $(P_1,P_2,P_3) \,\varepsilon\, \Sigma_K(P_{\mu\nu}(s))$ and $P_2 = DP$ where D is a diagonal matrix over K and P is a permutation matrix then P is of one of the forms

$$\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}$$

<u>Proof</u>: Either $P_2 : \underline{s} \to \underline{s}$ maps $s_i \to k_i s_i$ or $s_i \to k_i s_{2n-i}$ if P_1, P_3 exist such that $P_1 P_{\mu\nu}(P_2 s) = P_{\mu\nu}(s) P_3^{-1}$ and furthermore if one of these forms occurs for one i, it occurs for all i.

The utility of this analysis will emerge in the next section.

4. Realization of Characteristic Functions

We have defined the degree of $H(s,x,y)$ over K as the smallest integer n such that there exist vectors a_α, b_α, c_α, $\alpha = 1,\ldots n$ over K such that $H(s,x,y)= \sum_{\alpha=1}^{n} <a_\alpha,s><c_\alpha,x><b_\alpha,y>$. If $K = \mathbb{R}$ this corresponds to Strassen's definition of the rank of the tensor (h_{ijk}) ([22], p.10). We want to look at this from a different point of view. Define the matrices C, A(s) and B as

$$A(s) = \begin{bmatrix} <\underline{a}_1,s> & & & \\ & <\underline{a}_2,s> & & \\ & & \ddots & \\ & & & <\underline{a}_n,s> \end{bmatrix} B= \begin{bmatrix} \underline{b}_1' \\ \vdots \\ \underline{b}_n' \end{bmatrix}$$

$$C = [\underline{c}_1 \cdots \underline{c}_n]$$

It is clear that $G(s) = CA(s)B$ if and only if $H(s,x,y) = \langle x,G(s)y\rangle = \langle x,CA(s)By\rangle$. Any such triple of matrices (A,B,C) will be called a K-realization of G(s). The integer n is called the dimension of the realization. In view of the above construction we see that the following theorem holds.

Theorem 7: The K-degree of G(s) is equal to the smallest dimension of a K-realization of G(s).

Example 5: Consider a $\{1,0,-1\}$ -realization of 2nd degree polynomial multiplication. We multiply $(x_1+x_2z+x_3z^2)$ and $(y_1+y_2z+y_3z^2)$. The appropriate G(s) and a realization is

$$CA(s)B = \begin{bmatrix} s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \end{bmatrix}$$

where

$$C = B^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \text{and}$$

$$A(s) = \begin{bmatrix} s_1-s_2-s_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_3-s_2-s_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_5-s_3-s_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_2 \end{bmatrix}$$

Example 6: One form of the Strassen algorithm [21] for 2 by 2 matrix multiplication is revealed by

$$CA(s)B = \begin{bmatrix} s_1 & 0 & s_2 & 0 \\ 0 & s_1 & 0 & s_2 \\ s_3 & 0 & s_4 & 0 \\ 0 & s_3 & 0 & s_4 \end{bmatrix}$$

where

$$A(s) = \begin{bmatrix} s_1+s_4 & & & & & & \\ & s_2-s_1 & & & & & \\ & & s_2+s_4 & & & & \\ & & & s_3-s_4 & & & \\ & & & & s_3+s_1 & & \\ & & & & & s_1 & \\ & & & & & & s_4 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 & 1 & -1 & 0 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

This form has been studied by Fiduccia [6] and Gastinel [9] who extended the algorithm to obtain upper bounds on the difficulty of multiplying nxn matrices. We prove below that this factorization is minimal and study lower bounds on nxn matrix multiplication.

One advantage of working with realizations is that the application of standard algebraic machinery will allow one to describe in a convenient way some of the processes by which algorithms can be combined to form more complex ones. The three most basic processes of this type are sum, direct sum, and Kronecker product which are defined as follows:

i) $G(s)+H(t) = J(u) \overset{def}{=} \sum_{i=1}^{r+m} J_i u_i$.

$$\text{where } J_i u_i = \begin{cases} G_i s_i & 1 \leq i \leq m \\ H_{i-m} t_i & m < i \leq m+r \end{cases}$$

where m = dimension of s and r = dimension of t.

ii) $G(s) \oplus H(t) = J(u) \overset{def}{=} \begin{bmatrix} G(u_1) & 0 \\ 0 & H(u_2) \end{bmatrix}$

$u_1 = s$, $u_2 = t$, $u = u_1 \cup u_2$

iii) $G(s) \otimes H(t) = J(u) \overset{\text{def}}{=} \sum (G_i \otimes H_u) u_{ij}$

where $\otimes$ indicates Kronecker product of matrices and $u_{11} = s_1 t_1$, $u_{21} = s_2 t_1 \ldots u_{mr} = s_m t_r$ where m is the dimension of s and r is the dimension of t.

We have as an immediate consequence of these definitions the following theorem.

Theorem 8: If $G(s) = C_1 A_1(s) B_1$ and $H(s) = C_2 A_2(s) B_2$ then

i) $G(s)+H(t) = [C_1, C_2][A_1(s) \oplus A_2(t)] \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$

ii) $G(s) \oplus H(t) = (C_1 \oplus C_2)(A_1(s) \oplus A_2(t))(B_1 \oplus B_2)$

iii) $G(s) \otimes H(t) = (C_1 \otimes C_2)(A_1(s) \otimes A_2(t))(B_1 \otimes B_2)$

where $s_i t_j$ are to be interpreted as distinct indeterminants for distinct index pairs (i,j).

Corollary:

i) $\delta_K(G(s)+H(t)) \leq \delta_K(G(s))+\delta_K(H(t))$

ii) $\delta_K(G(s) \oplus H(t)) \leq \delta_K(G(s)) + \delta_K(H(t))$

iii) $\delta_K(G(s) \otimes H(t)) \leq \delta_K(G(s))\delta_K(H(t))$

Remarks:

i) If $K = \mathbb{Z}$ (the integers) or $\mathbb{R}$, then it is easily verified that $P_n(s) \otimes P_m(t) \underset{K}{\supseteq} P_{nm}(U)$ and therefore $\delta_K(P_{2^n}(s)) \leq [\delta_K(P_2(s))]^n = 3^n$. We will discuss this inequality further in a later section.

ii) If $K = \mathbb{Z}$ or $\mathbb{R}$, then it may be verified

$$M_{\mu_1,\nu_1,\omega_1}(s) \otimes M_{\mu_2,\nu_2,\omega_2}(t) \underset{K}{\sim} M_{\mu_1\mu_2,\nu_1\nu_2,\omega_1\omega_2}(u)$$

and thus $\delta_z(M_{2^n}(s)) \leq 7^n$.

For example ii) gives the upper bound 168 on 6 by 6 matrix multiplication by viewing this as the Kronecker product of $M_{222} \otimes M_{333}$ but gives a bound of 165 by viewing it as $M_{2,3,2} \otimes M_{3,2,3}$.

As is apparent from the work of Hopcroft, Kerr, and Musinski [11], [12] it is useful to know not just one realization of a given G(s) but rather all realizations of it. For example, their algorithm for multiplying 2 by 3 matrices by 3 by 3 matrices uses 3 different forms of 2 by 2 matrix multiplication in a subtle way. The concept of a realization is ideal for describing how one gets large classes of algorithms from a given one, as we will now describe.

If $G(s) = CA(s)B$ and if $P_1G(P_2s)P_3 = G(s)$ then it may be verified that $G(s) = C_1A_1(s)B_1$ where

$$C_1 = P_1CE_1$$
$$B_1 = E_2BP_3$$
$$A_1 = E_1^{-1}E_2^{-1}AP_2$$

and E_i are any diagonal matrices. This comes about by inserting $E_1E_1^{-1}$ and $E_2E_2^{-1}$ in the expression for G(s) according to

$$CA(s)B = CE_1E_1^{-1}A(s)E_2^{-1}E_2B$$

the matrix A and A(s) being related as indicated at the start of this section. If we have in mind realization over a subring $K \subset \mathbb{C}$ then of course the elements of E_i and P_i must come from this ring. It would be especially pleasant if it turned out that all minimum dimensional realizations were so related. We therefore pose an open problem.

Open Problem: Are all K-realizations of G(s) of minimal degree generated from a given one by considering permutation equivalences and the action of $\Sigma_K(G(s))$ according to the above equations?

If the answer to this question is true, as we conjecture it is, then it will be possible to apply the results of the previous section and the group structure of $\Sigma_K(G(s))$ to greatly reduce the effort needed to determine, for a given n, if there exists a dimension n realization of G(s) over K. Furthermore, if all realizations of smaller problems are known, this result can be used to combine them, as done by Hopcroft and Musinski [12], to form realizations of large problems. If $K = \mathbb{Z}$, then all K-realizations of minimal degree are related in this manner for complex multiplication (shown in Appendix A1) and 2x2 matrix multiplication (implicit in Hopcroft-Musinski [12]).

5. Effects of Preconditioning and Field Extensions

Most of the results presented here regarding lower and upper bounds on the evaluation of bilinear and n-linear operations are independent of the choice of the set K. This is clearly desirable in the development of a general theory and to the derivation of upper and lower bounds on the operations which we study. However, in practice, the nature of the set K is very often of central importance. For example, Winograd [28] shows that while the minimal algorithm for $P_{3,3}$ over the real or complex field is of degree 5, it always requires that many divisions. The minimal algorithm over the set {1,0,-1} is of degree 6 but of a very simple form and desirable in actual practice. However, it appears that for large enough values of n, minimal degree algorithms for $P_{n,n}$ over $\mathbb{R}$ or $\mathbb{C}$ require less work than realization of minimal degree over the integers. Before defining a cost criterion which will enable us to measure the relative efficiencies of algorithms in a more complete way, we will study the effect of field extensions on a bilinear operation's degree in order to understand what we want to include in a cost criterion. Our notation is standard and can be found in Lang [16].

Theorem 9: i) If G(s) is any characteristic function over $\mathbb{Z}$, then

$$\delta_{\mathbb{Z}}(G(s)) \geq \delta_{\mathbb{Q}}(G(s)) \geq \delta_{\mathbb{R}}(G(s)) \geq \delta_{\mathbb{C}}(G(s))$$

ii) If K_1 and K_o are subfields of $\mathbb{C}$ such that K_1 is a normal extension of K_o of even degree, then there exists a characteristic function G(s) over K_o such that

$$\delta_{K_o}(G(s)) > \delta_{K_1}(G(s))$$

Furthermore, for any integer p there is a characteristic function $G_p(s)$ such that

$$\delta_{K_o}(G_p(s)) > \delta_{K_1}(G_p(s))+p$$

iii) If K_2 is a normal extension field of K_1 which is non-trivial then there exists a characteristic function G(s) over K_1 such that

$$\delta_{K_2}(G(s)) < \delta_{K_1}(G(s))$$

<u>Proof</u>: i) It is clear that if $K_2 \supset K_1$ then any K_1-realization is also a K_2-realization.

ii) If K_1 is a normal extension of K_o of even degree, then there exists $k \in K_o$ such that $\sqrt{k} \notin K_o$ and the extension can be factored through the field $K_o(\sqrt{k})$. It suffices to prove that there is a characteristic function, G(s), such that $\delta_{K_o(\sqrt{k})}(G(s)) < \delta_{K_o}(G(s))$. Such a characteristic function is given by $G(s) = \begin{bmatrix} s_1 & s_2 \\ s_2 & ks_1 \end{bmatrix}$ and a p-fold direct sum of this characteristic function yields one of degree p less over $K_o(\sqrt{k})$ than over K_o.

iii) For a polynomial $\phi(x)$, let $G_\phi(s)$ be the characteristic function corresponding to multiplying polynomials of degree less than the degree of $\phi(x)$ modulo the constraint $\phi(x) = 0$. Corresponding to any non-trivial normal extension K_2 of K_1 is a polynomial $r_{K_2,K_1}(x)$ irreducible over K_1 which has K_2 as its splitting field, and it is easily verified that $\delta_{K_2}(G_{r_{K_2,K_1}}(s)) < \delta_{K_1}(G_{r_{K_2,K_1}}(s))$. Fiduccia's discussion of companion matrices [6] also considers this operation.

<u>Remark:</u> It is worthwhile to note that all degree 2 factorizations of $G_o(s)$ are of the form $G(s) = (CD_1)(D_1^{-1}A(s)D_2^{-1})(D_2B)$ where

$C = [\frac{1}{\sqrt{k}}\ \ \frac{1}{-\sqrt{k}}]$, $A(s) = \begin{bmatrix} \frac{s_1+s_2}{2} & \\ & \frac{s_1-s_2}{2\sqrt{k}} \end{bmatrix}$, $B = C^T$ and D_1 and D_2 are nonsingular diagonal matrices.The degree 3 factorization $G(s)=KH(s)F$

$F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = K^T$, $H(s) = \begin{bmatrix} s_1-s_2 & & \\ & ks_1-s_2 & \\ & & s_2 \end{bmatrix}$ is more desirable for computation than any degree 2 factorization and we wish to design a cost criterion which reflects this difference.

It is also helpful to understand how much a field extension can help in order to find a cost criterion which is as realistic as possible.

Theorem 10: i) If $G(s)$ is a characteristic function over $\mathbb{Z}$, then there exists an integer N such that $\delta_{\mathbb{Z}}(NG(s)) = \delta_{\mathbb{Q}}(G(s))$

ii) For any $G(s)$ over $\mathbb{R}$, $\delta_{\mathbb{C}}(G(s)) \geq \frac{1}{3}\ \delta_{\mathbb{R}}(G(s))$.

Proof: i) If $G(s) = CA(s)B$ over Q, then we can find integers N_1, N_2, and N_3 such that N_1C, $N_2A(s)$, N_3B are matrices over $\mathbb{Z}$. If we take $N = N_1N_2N_3$, then $NG(s) = (N_1C)(N_2A(s))N_3B$ and therefore $\delta_{\mathbb{Z}}(NG(s)) = \delta_{\mathbb{Q}}(G(s))$.

ii) If $G(s) = CA(s)B$ over $\mathbb{C}$, then we can write $C = C_R+iC_I$, $A(s) = A_R(s)+iA_I(s)$, $B = B_R+iB_I$ where C_R, C_I, A_R, A_I, B_R and B_I are real matrices, $G(s) = C_R(A_R(s)B_R-A_I(s)B_I)-C_I(A_I(s)B_R+A_R(s)B_I)$, and $C_R(A_I(s)B_R+A_R(s)B_I)+C_I(A_R(s)B_R-A_I(s)B_I) = 0$. In this case,

$$G(s) = [C_R+C_I\ \ \ C_R-C_I\ \ \ -2C_I] \begin{bmatrix} \frac{A_R(s)+A_I(s)}{2} & & \\ & \frac{A_R(s)-A_I(s)}{2} & \\ & & A_I(s) \end{bmatrix} \begin{bmatrix} B_R-B_I \\ B_R+B_I \\ B_R \end{bmatrix}$$

is a realization of thrice the size over $\mathbb{R}$.

Remark: In light of i) in Theorem 10, it seems most realistic to have in our cost criterion allowance for scaling, since it is entirely possible that there exist characteristic functions of much larger degree over $\mathbb{Z}$ than $\mathbb{Q}$ which can be scaled to yield characteristic functions of the same degree.

For a realization CA(s)B of G(s), we will define the work of the algorithm by $w_k(C,A,B) = k[d(C)+d(A)+d(B)]+(\text{dimension }(A))$ where d(X) is the difficulty of multiplying the matrix X by an arbitrary vector and k is a given weight. Furthermore, we shall define the cost c_k of a given characteristic function with respect to the weight k through the following

Definition:
$$c_k^1(G(s)) \triangleq \min\{w_k(C,A,B)\,|\,CA(s)B=G(s)\}$$
$$c_k^2(G(s)) \triangleq \min\{2w_{k/2}(C,A,B)\,|\,CA(s)B = NG(s) \text{ for N any nonzero scalar}\}$$
$$c_k(G(s)) \triangleq \min(c_k^1(G(s)),c_k^2(G(s))).$$

This definition was chosen because it provides a realistic complexity measure and is general enough to handle the situations described in Theorem 9 and 10. In much of what follows, we shall let k = 0 and seek only minimum degree algorithms. Regardless of our choice of k, the following will characterize bounds on the complexity of evaluating a set of bilinear forms.

Theorem 11: i) If $\delta_{\mathbb{C}}(G(s)) = m$, then no algorithm for evaluating the set of bilinear forms in G(s) requires less work than the computation of m products of linear forms in the inputs.

ii) If $F_3 = \{1, 0, -1\}$ and $\delta_{F_3}(G(s)) = n$, then the minimal algorithm for evaluating the set of bilinear forms in G(s) requires less work than the computation of n products of linear forms in the inputs.

When considering a family of characteristic functions, we will write $c(\mathscr{F}_n(s)) = O(f(n))$ if for all nonzero k, $c_k(\mathscr{F}_n(s))$ grows as $O(f(n))$.

CHAPTER III

UPPER AND LOWER BOUNDS

0. Introduction

In the preceeding chapter, the general bilinear multiplication problem was described. It is clear from that description, that the major problem to be studied is the determination of methods for evaluating $\delta_K(G(s))$ and $c_k(G(s))$ for arbitrary characteristic functions. Two classes of methods exist for evaluating these functions, analysis methods and synthesis methods. Analysis methods are used for determining lower bounds on $\delta_K(G(s))$ and $c_k(G(s))$ by analyzing the structure of G(s) and determining the difficulty of realizing characteristic functions having this structure. Synthesis methods consist of finding algorithms for generating realizations of arbitrary characteristic functions. Three types of analysis procedures and a single synthesis procedure are discussed here.

The first analysis procedure consists of extending the results of Winograd [26] and Fiduccia [6] on the application of linear dependence to the determination of upper and lower bounds on matrix-vector products to the present framework. These methods are used to determine lower bounds on the partial evaluation of third order tensors at a pair of vectors. This procedure is most valuable in finding lower bounds on $\delta_K(G(s))$. Next, methods of relating $c_k(P_{n,m}(s))$ to $c_k(M_{p,q,r}(s))$ are discussed for various values of n,m,p,q, and r. These results are useful in determining lower bounds on the work necessary to compute with a family of characteristic functions. Our final analysis procedure consists of relating the determination of lower bounds on $\delta_{\mathbb{Z}}(G(s))$ to some problems in algebraic coding theory. The synthesis procedure we discuss is one that finds realizations of G(s) by finding factorizations of the structural tensor (h_{ijk}).

1. Lower Bounds Obtained by Partitioning

Winograd [26] and Fiduccia [6] have shown how linear dependence can be exploited in the analysis of the arithmetic complexity of multiplying a matrix times a vector. To fully utilize linear dependence arguments, it is necessary to deal with linear dependence as applied to rank one matrices. The problems are highly non-trivial because rank one matrices do not form a vector space. One aspect of the Winograd argument is the fact that rank is invariant under interchange of indices, i.e., transposition, for a second-order tensor (i.e. a matrix). This yields bounds for multiplying particular matrices by vectors. By comparison, what we do here is to find lower bounds for the partial evaluation of certain third-order tensors at pairs of vectors. Here it is again true that the rank of the tensor is invariant under permutation of indices and the six permutations which exist can be used to provide insight into the difficulty of this evaluation. By focusing attention on the structural constants (h_{ijk}) rather than the $\{G_i\}$ we are able to take maximum advantage of the flexibility provided by the invariance of degree under interchanges of the roles of the row, column and indeterminant variables.

We begin with two lemmas which form the basis for a number of new results. In this section s^i refers to an ordered subset of the set of indeterminants s whereas s_i is, as always, a particular element of s.

Lemma: If $G(s) = [G_1(s^1) | G_2(s^2)]$ where s^1 and s^2 are nonoverlapping sets of indeterminates, then

$$\delta[G(s)] \geq \delta[G_1(s^1)] + \text{index of } (G_2(s^2))$$

Proof: Let G_1 be expressed as

$$G_i = \sum_{j=1}^{n} a_{ij} D_j$$

and let $\tilde{D}_j$ denote the restriction of D_j to the left part setting the right part to zero. Then clearly at most $\nu = n-\delta[G_1(s^1)]$ linearly independent relations of the form

$$\sum_{j=1}^{n} \beta_{ij}\tilde{D}_j = 0 \qquad i=1,\ldots\nu$$

can exist. For each such relation, we can define a matrix

$$A_i = \sum_{j=1}^{n} \beta_{ij}(D_j-\tilde{D}_j)$$

such that the set of matrices $A_1,\ldots A_\nu$ form a basis for the subspace of the linear span of the D_j consisting of matrices vanishing on the left. These matrices are of arbitrary rank, but at least index $[G_2(s^2))]$ are required to realize $G_2(s^2)$ and therefore $\nu \geq$ index $[G_2(s^2)]$. Similarly $n \geq \delta[G_2(s^2)]$+index $(G_1(s^1))$.

<u>Lemma</u>: If

$$G(s) = \begin{bmatrix} G_1(s) & 0 \\ 0 & 0 \end{bmatrix} ,$$

then all minimal realizations of $G(s)$ are of the form

$$G(s) = \begin{bmatrix} C \\ 0 \end{bmatrix} [A(s)][B \quad 0] \qquad \text{where } CA(s)B = G_1(s)$$

<u>Proof</u>: It is enough to establish the lemma in the special case if

$$G(s) = [G_1(s),0]$$

where 0 indicates a single column since the more general case follows by adding more columns of zeros one at a time and by transposition.

If (G(s),0) has ann dimensional realization with any nonzero entries in the last column then we will show that the degree of G(s) is n-1 or less. Suppose

$$(G(s),0) = \sum_{i=1}^{m} (G_i,0)s_i = \sum_{i=1}^{m} \sum_{j=1}^{n} s_i\alpha_{ij}D_j$$

$$= \sum_{i=1}^{m} s_i\Big(\sum_{j=1}^{\nu}\alpha_{ij}(b_jc_j,b_j)+\sum_{j=\nu+1}^{n}\alpha_{ij}(b_jc_j,0)\Big)$$

where we have used the fact that we can normalize the last column of the dyed, if it is nonzero, by appropriate choice of the row vectors c_j. Now since .

$$\sum_{j=1}^{\nu} \alpha_{ij}b_j = 0$$

We can subtract this sum, post multiplied by $(c_1,1)$, obtaining

$$(G(s),0) = \sum_{i=1}^{m} s_i \sum_{j=1}^{n} \alpha_{ij}(b_j(c_j-c_1),0)$$

$$(G(s),0) = \sum_{i=1}^{m} s_i\Big(\sum_{j=2}^{\nu} \alpha_{ij}(b_jc_j,0) + \sum_{n=\nu+1}^{n} \alpha_{ij}(b_jc_j,0)\Big)$$

which shows that G(s) is of degree n-1 or less.

We define the row (resp. column) rank of G(s) as the number of K-linearly independent rows (resp. columns) of G(s) and note that row rank (G(s)) $\neq$ column rank (G(s)) in general. For example, if G(s) = $[s_1,s_2]$, then the row rank is 1 and the column rank 2. Note further that if the G_i are linearly independent then the index of G(s) is equal to the column rank of $\check{G}(x)$, the row rank of G(s) is equal to the index of $\check{G}(x)$, and the column rank of G(s) is equal to the row rank of $\check{G}(x)$.

Theorem 1: Suppose that $\{G_i\}$ are linearly independent and $G(s)$ is of full row and column rank, then

i) if $G(s) = [G_1(s^1) | G_2(s^2)]$ where s^1 and s^2 are nonoverlapping sets of indeterminates then $\delta[G(s)] \geq \delta[G_1(s^1)]+\max(\text{column rank }(G_2(s^2)), \text{index }(G_2(s^2)))$.

ii) if $G(s) = \begin{bmatrix} G_1(s^1) \\ G_2(s^2) \end{bmatrix}$ where s^1 and s^2 are nonoverlapping sets of indeterminates, then

$$\delta[G(s)] \geq \delta[G_1(s^1)]+\max(\text{row rank }(G_2(s^2)), \text{ index }(G_2(s^2)))$$

iii) if $G(s) = \begin{bmatrix} G_1(s) & 0 \\ 0 & G_2(s) \end{bmatrix}$

then

$$\delta[G(s)] \geq \delta[G_1(s)]+\max(\text{column rank}(G_2(s)), \text{ row rank }(G_2(s)))$$

iv) The degree of

$$G(s) = \begin{bmatrix} G_1(s^1) & 0 \\ 0 & G_2(s^2) \end{bmatrix}$$

equals the sum of the degrees of $G_1(s^1)$ and $G_2(s^2)$ if s^1 and s^2 are nonoverlapping sets.

Proof: i) is true by the previous two lemmas and ii) and iii) are equivalent to i) under the action of the permutation equivalence group Γ. To prove iv), we observe that the number of rank one matrices which intersect the upper left corner defined by $G_1(s^1)$ equals or exceeds $\delta[G_1(s^1)]$, and the number which intersect the lower left corner equals or exceeds $\delta[G_1(s^2)]$. There cannot be linear relations between these because there are none between the elements of s^1 and s^2. Thus there are $\delta[G_1(s^1)]+\delta[G_2(s^2)]$ rank one matrices which intersect the upper left or lower right corner. This shows that the degree is at least $\delta[G_1(s^1)]+\delta[G_2(s^2)]$ but clearly it is not higher.

Corollary: If

$$G(s) = \begin{bmatrix} G_1(s) & 0 \\ 0 & \begin{matrix} G_2(s^1) \\ G_3(s^2) \end{matrix} \end{bmatrix}$$

where s^1 and s^2 are nonoverlapping setsof indeterminates in s, then

$$\delta[G(s)] \geqslant \delta \begin{bmatrix} G_1(s) & 0 \\ 0 & S_2(s^1) \end{bmatrix} + \text{row rank } [G_3(s^2)]$$

Proof: Trivial by extension of the preceeding theorem.

We now use these results to obtain a new lower bound on matrix multiplication.

Lemma: $\delta(M_{\mu,\nu,\omega}(s)) \geqslant \delta(M_{\mu,\nu,\omega-1}(s))+\mu+\nu-1$

Proof: Observe that in the notation of Chapter 2

$$M_{\mu\nu\omega}(s) = \begin{bmatrix} M_{\mu,\nu,\omega-1}(s) & 0 \\ 0 & \theta_{\mu,\nu}(s) \end{bmatrix}$$

and apply the corollary.

Theorem 2 : $\delta(M_{\mu,\nu,\omega}(s)) \geqslant (\mu\nu+\mu\omega+\nu\omega)-(\mu+\nu+\omega)+1$

Proof: $\delta(M_{\mu,\nu,\omega}(s)) \geqslant (\mu+\nu-1)(\omega-1)+\delta(M_{\mu,\nu,1}(s))$ by applying the recursion in the preceeding lemma. By using the identity $\delta(M_{\mu,\nu,\ 1}(s)) = \delta(\theta_{\mu,\nu}(s)) = \mu\nu$ we obtain the bound given.

Observe that these results establish, for example, that 2 by 2 matrix multiplication requires 7 scalar multiplication regardless of the set K.

Previous proofs of this result by Hopcroft-Kerr [11](for $K = \mathbb{Z}$) and Winograd [27] (for $K = \mathbb{C}$) involve more detailed computation than the proof given here. On the other hand, Hopcroft and Kerr [11]have shown that multiplication of a 2x2 matrix by a 2xn matrix can be done in $\lceil \frac{7n}{2} \rceil$ multiplications and no scheme requiring fewer multiplications is possible if one considers integer realizations only. The above inequalities do not yield this bound, but instead show that at least $3n+1$ multiplications are necessary for any set K. Using the Hopcroft-Kerr result in our recursion gives the following lower bound on integer realizations.

<u>Corollary</u>: $\delta_{\mathbb{Z}}[M_{\mu\nu\omega}(s)] \geq (\mu\nu+\mu\omega+\mu\omega)-(\nu+\omega+\lfloor \frac{\mu}{2} \rfloor)$.

In particular this shows that over $\mathbb{Z}$, 3 by 3 matrix multiplication requires 20 scalar multiplications -- improving by 2 the best previously known bound. A summary of best known upper and lower bounds on $\delta_{\mathbb{Z}}(M_{p,q,r}(s))$ and $\delta_{\mathbb{C}}(M_{p,q,r}(s))$ for various p, q and r is given in Appendix A.4.

It is interesting to note that the recursion yielding Theorem 2 also gives rise to the inequalities

$$\delta(M_{n-1}(s))+3n^2-3n+1 \geq \delta(M_n(s)) \geq \delta(M_{n-1}(s))+6(n-1)$$

and therefore, if for some i, $\delta(M_i) > 3i^2-3i+1$, then for all $n > i$, it is also true that $\delta(M_n) > 3n^2-3n+1$. In particular, if it is shown that $\delta(M_3)$ is greater than 19, 19 being the bound implicit in the above, then the bound on $\delta(M_n)$ will be sharpened, for all $n > 3$.

We can further extend the result of Theorem 2 to the following recursion.

Theorem 3: If there exists an integer p and a number $f(p)$ such that for all $n > f(p)$, $\delta(M_{n,n,p}(s)) \geq \alpha n^2+\beta n+\gamma$, then for all $n > f(p)$ $\delta(M_{n,n,n}(s)) \geq (\alpha+2)n^2+(\beta-2p-1)n+(\gamma+p)$.

Proof: By applying the recursion in the Lemma preceeding Theorem 2, the result is obvious.

The two recursions given here are valuable in the interpretation and extension of any new lower bounds on matrix multiplication. The bound of $3n^2-3n+1$ given above merely results from applying Theorem 3 to the lower bound of n^2 on $\delta(M_{n,n,1})$, matrix-vector product, given by Winograd [26]. If the upper bound of $\frac{3n^2}{2} + 0(n)$ on $\delta(M_{n,n,2}(s))$ given by Hopcroft and Kerr [11] is shown to be a lower bound also, then the result of Theorem 3 would be to extend $\delta(M_{n,n,n}(s))$ to $3\frac{1}{2}n^2+0(n)$. A detailed analysis of this and other cases leads to strong support for the following conjecture which occupies a central role in the minds of most complexity theory researchers.

Conjecture: $\delta(M_{n,n,n}(s))$ grows asymptotically faster than $0(n^2)$.

2. Lower Bounds Obtained by Reduction

In the last section, a new lower bound on the complexity of matrix multiplication was obtained by studying the structure of $M_{\mu,\nu,\omega}(s)$. The result was a lower bound on this operation which could be stated in closed form. Our goal in the present section will be to obtain lower bounds on $M_{\mu,\nu,\omega}(s)$ in terms of lower bounds on other operations. What we seek to do is to provide a framework in which polynomial multiplication and matrix multiplication problems can be studied togethwe so that advances in lower bounds on $P_{u,v}(s)$ can be used to yield new lower bounds on $M_{\mu,\nu,\omega}(s)$ and new upper bounds for $M_{\mu,\nu,\omega}(s)$ can be used to generate new upper bounds for $P_{u,v}(s)$. In particular, if the conjectured lower bound of $0(n \log n)$ on $c(P_{n,n}(s))$ is proven true, then the lower bound on $M_{n,n,n}$ can be extended. We begin by studying the operation $M_{p,q,r} \otimes P_{u,v}$ and observe that

Lemma: If γ is a member of the permutation equivalence group Γ, then $\gamma(A(s) \otimes B(t)) = ((\gamma A)(s) \otimes (\gamma B)(t))$.

Proof: If (h_{ijk}) and (ℓ_{pqr}) are third order tensors and $\pi \in S_3$, the permutation group on three letters, we wish to show that if (m_{stv}) is the stv element of $h \otimes \ell$, then $m_{\pi(s),\pi(t),\pi(v)} = (h_{\pi(i),\pi(j),\pi(k)}) \otimes (\ell_{\pi(p),\pi(q),\pi(r)})$. In this formulation the result is obvious.

Corollary: $M_{p,q,r}(s) \otimes P_{u,v}(t)$ is permutation equivalent to $M_{r,p,q}(s) \otimes T_{v,u,}(t)$.

By examining the form of this latter operation, the following result is obvious.

Theorem 4: $(M_{p,q,r} \otimes P_{u,v})(s) \subseteq M_{rv,pu,q}(t)$

In its present form, Theorem 4 is very applicable to the study of computing products of matrices of polynomials or equivalently, polynomials whose coefficients are matrices. By making the substitution p=r=1, we can derive an interesting lower bound $(M_{1,q,1} \otimes P_{u,v}(s)) \subseteq M_{v,u,q}(t))$ on $M_{v,u,q}(s)$. This implies that the mültiplication of a vxu matrix by a uxq matrix is a harder operation than computing the inner product of a q-tuple of u-1 degree polynomials with a q-tuple of v-1 degree polynomials. Further examination of lower bounds on the operation $(M_{1,q,1} \otimes P_{u,v})(s)$ may yield improved lower bounds on $\delta(M_{n,n,n}(s))$ in various contexts. For example, if u = 1 (or equivalently, if v=1), then it is well known (see e.g. Winograd [26] that $\delta(M_{1,q,1} \otimes P_{1,v}) = \delta(M_{1,q,1}) \otimes \delta(P_{1,v}) = qv = \delta(M_{v,1,q})$. If we could extend this result to the case where u = 2, then it would be possible to show that $\delta_{\mathbb{Z}}(M_{v,2,q}) \geq q \left\lceil \frac{3v}{2} \right\rceil$ which almost reaches the upper bound of $\left\lceil \frac{3vq+\max(v,q)}{2} \right\rceil$ given by Hopcroft and Kerr [11].

Another relationship between polynomial multiplication and matrix multiplication can be established by examining the internal structure of matrix multiplication. If $A = (a_{ij})$ and $B = (b_{ij})$ are pxq and qxr matrices respectively, then $M_{p,q,r}(s)$ is the operation of computing the product AB. This operation is dominated by the operation of computing AC where C is a Toeplitz matrix of the form

$$C = \begin{bmatrix} b_{11} & b_{12} & & b_{1,r-q+1} & 0 & \cdots & 0 \\ 0 & b_{11} & & & b_{1,r-q+1} & & \\ & 0 & & & & & \\ & \vdots & & & & & 0 \\ 0 & 0 & b_{11} & & & & b_{1,r-q+1} \end{bmatrix}$$

Now, computing the product AC is identical to computing the set of polynomial products, $a_1(x)b(x)$, $a_2(x)b(x),\ldots a_p(x)b(x)$, for $a_i(x) = \sum_{j=1}^{q} a_{ij}x^{j-1}$ and $b(x) = \sum_{i=1}^{r-q+1} b_{1i}x^{i-1}$, as can be verified by direct expansion. We will let $P^p_{q,r-q+1}(s)$ represent the characteristic function for this operation. While the relationship between $P^p_{q,r-q+1}(s)$ and $\bigoplus_{i=1}^{p} P_{q,r-q+1}(t^i)$ is not known, it is possible to establish the dominance relation.

Lemma: $P^p_{q,r-q+1}(s) \supseteq P_{pq,r-q+1}(t)$.

Proof: In terms of the exposition above, any computation of $\{a_i(x)b(x) | i=1,\ldots p\}$ can be easily transformed to a computation of $\alpha(x)b(x)$ where $\alpha(x) = \sum_{i=1}^{p} a_i(x)x^{(i-1)q}$ and thus $P^p_{q,r-q+1}(s) \supseteq P_{pq,r-q+1}(t)$. The converse may not be true but is unnecessary to what follows.

We can combine these facts to prove the next result

Theorem 5: $M_{p,q,r}(s) \supseteq P_{pq,r-q+1}(t)$

Proof: As shown above $M_{p,q,r}(s) \supseteq P_{pq,r-q+1}(u) \supseteq P_{pq,r-q+1}(t)$ and transitivity is a property of $\supseteq$.

As an application of this theorem, we obtain the following results

Corollary: $M_{n,n,n}(s) \supseteq \frac{1}{2} P_{n^2,n}(t)$

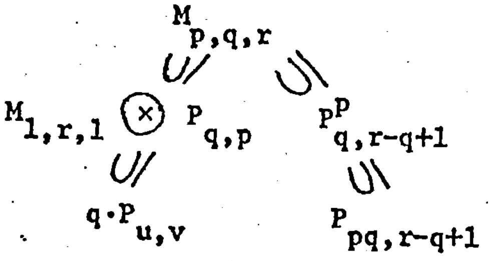
Proof: It is easy to see that $2M_{n,n,n}(s) \supseteq M_{2n-1,n,n}(u)$ and $M_{2n-1,n,n}(u) \supseteq P_{n^2,n}(t)$ by Theorem 5.

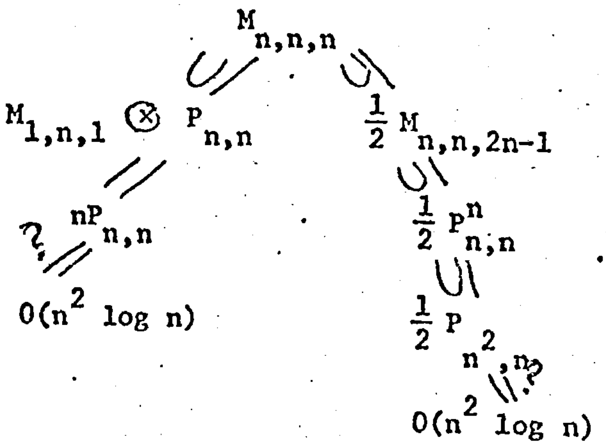
This result and the widely held conjecture that $c(P_{n,n})$ grows as $O(n \log n)$ lend strong support to

<u>Conjecture</u>: $c(P_{n^2,n})$ grows as fast as $O(n^2 \log n)$ and therefore $c(M_{n,n,n})$ grows at least that fast.

As we have mentioned before, verification of $c(M_{n,n,n}) \geq O(n^2 \log n)$ is a pivotal step in arithmetic complexity as it is now conceived.

The results of this section can be summarized by the figures below





3. Lower Bounds for Realizations over the Integers

Thus far the methods discussed here have been completely field independent. Moreover, a particular choice of the set K has not been necessary. In this section, we specialize to realizations over the integers. The approach used here exploits the fact that every realization over the integers yields a realization over $\mathbb{Z}_2 = \{0,1\}$ simply by replacing all entries in A, B, and C by their residues mod 2. This device has been used effectively by Hopcroft et al. [11]-[12]. This approach is useful if all elements in (h_{ijk}) are 1 or 0, in which case we say that the degree of G(s) over $\mathbb{Z}_2$ (i.e. $\delta_{\mathbb{Z}_2}[G(s)]$) is the dimension of the smallest realization (A,B,C) such that $G(s) \equiv CA(s)B(\text{mod } 2)$. In this section we only consider (0,1) valued structural constants. It is clear that for these characteristics functions, $\delta_{\mathbb{Z}}[G(s)] \geq \delta_{\mathbb{Z}_2}[G(s)]$. Using this approach, it is possible to make contact with results on the structure of 0-1 matrices discovered by algebraic coding theorists.

We adopt the following notation for characterizing the set of all possible characteristic functions and all matrices contained in G(s).

Let $G(s) = \sum_{i=1}^{m} s_i G_i$ and $0 < k < 2^m$ be an integer with binary expansion $k = k_1 k_2 \cdots k_m$ then write

i) $G(s)\Big|_k \overset{\text{def}}{=} \left(\sum_{i=1}^{m} G_i k_i\right)_{\text{mod } 2}$

ii) $G(s)\Big|^k \overset{\text{def}}{=} \sum_{i=1}^{m} G_i k_i s_i$

We also define the special notation $\bigcup_{i=1}^{m} v_i$ where v_i are $\{0,1\}$ vectors to be a vector which has a one in the kth entry if any of the v_i do and is otherwise zero in the kth entry.

Theorem 6 : If $\delta_{2}(G(s)) = n$ then there exist m vectors $v_1, \ldots v_m \varepsilon \{0,1\}^n$ such that

$$\text{i)} \quad \text{rank}_{\text{mod } 2}(G(s)|_k) \leq |(\sum_{i=1}^{m} v_i k_i)_{\text{mod } 2}|$$

$$\text{ii)} \quad \delta_{2}(G(s)|^k) \leq |\bigcup_{i=1}^{m} v_i k_i|$$

for all k such that $0 < k < 2^m$ where $|.|$ represents the Hamming norm.

Proof: If $\delta_{\mathbb{Z}_2}[G(s)] = n$, and $G(s) = CA(s)B$ where the iith entry of $A(s)$ is $\langle a_i, s\rangle$, then form the matrix A_I such that the ith column of A_I is a_i.

i) then $G_i = \sum_{j=1}^{n} a_{ij} D_j (\text{mod } 2)$ and

$$G(s)|_k = \sum_{i=1}^{m} G_i k_i = \sum_{i=1}^{m} \sum_{j=1}^{n} k_i a_{ij} D_j = \sum_{j=1}^{n} D_j \sum_{i=1}^{m} k_i a_{ij}$$

and therefore $\text{rank}\ \mathbb{Z}_2(G(s))|_k \leq |\sum_{i=1}^{m} k_i a_i|$.

$$\text{ii)} \quad G(s)|^k = \sum_{i=1}^{m} G_i k_i s_i = \sum_{i=1}^{m} \sum_{j=1}^{n} k_i a_{ij} s_i D_j = \sum_{i=1}^{n} D_j \sum_{i=1}^{m} k_i a_{ij} s_i$$

and by definition of $\delta_{\mathbb{Z}_2}(G(s)|^k)$, at least one of $\{k_i a_{ij}\}_{i=1}^{m}$ must be nonzero for at least $\delta_{\mathbb{Z}_2}(G(s)|^k)$ values of j.

There are results in algebraic coding theory which are particularly useful in computing lower bounds for $\delta_{\mathbb{Z}}(P_{n,m}(s))$ and of particular interest is the following function which has been defined by Hamming and tabulated (for small d,k) by Calabi and Myrvaagnes [4].

Let $N(k,d)$ be the least n such that there exist k vectors in $\{0,1\}^n$ such that all words in their linear span have Hamming weight of at least d.

Greismer has shown that $N(k,d)$ satisfies the recursion $N(k,d) \geq N(k-1, \left\lfloor \frac{d+1}{2} \right\rfloor)+d$.

Theorem 7 : $\delta_{\mathbb{Z}_2}(P_{n,m}(s)) \geqslant \max(N(n,m),N(m,n))$.

Proof: Define the n by n+m-1 characteristic function

$$T_{n,m}(s)= \begin{bmatrix} s_1 & s_2 & s_3 & \cdots & & \cdots & s_m & 0 & \cdots & 0 \\ 0 & s_1 & s_2 & \cdots & & \cdots & & s_m & \cdots & 0 \\ 0 & 0 & s_1 & \cdots & & \cdots & & & \cdots & \\ \vdots & & & & & & & & & \\ 0 & 0 & 0 & \cdots & s_1 & s_2 & \cdots & \cdots & \cdots & s_m \end{bmatrix}$$

corresponding to multiplication of an mxn Toeplitz matrix by an n-vector, then $P_{n,m}(s)$ is permutation-equivalent to both $T_{n,m}(t)$ and $T_{m,n}(u)$. It is clear from the first part of Theorem 11 that $\delta_{\mathbb{Z}_2}(T_{n,m}(t)) \geqslant N(n,m)$. Thus $\delta_{\mathbb{Z}_2}(P_{n,m}(s)) = \delta_{\mathbb{Z}_2}(T_{n,m}(t)) = \delta_{\mathbb{Z}_2}(T_{m,n}(u)) \geqslant \max(N,(m,n),N(n,m))$.

Remark: This theorem and other observations yield exact bounds if n=2 or m=2 or n=m<6. These results along with realization are presented in Appendix A.2.

4. Upper Bounds by Tensor Ranks

As we noted previously, a bilinear multiplication operation gives rise to a set of structural constants (h_{ijk}) and any algorithm for realizing the operation consists of a factorization of the form (6). As Strassen [22] observed, the degree of the minimal algorithm for computing a given operation is equal to the minimum value of n for which the sum (6) holds which is equal to the rank of the tensor with ijk component given by h_{ijk}. Finding the rank of a third order tensor is a non-trivial problem and at present no extension of the standard tools (e.g. Gaussian elimination) or canonical forms (e.g. Jordan canonical form) for ranking second order tensors (i.e. matrices) exists. This problem received much attention during the 1920's and 1930's (see e.g. Hitchcock [10] and Oldenburger [18]) and as early as the 1890's, Kronecker et al (see Gantmacher [8]) were studying methods of ranking tensors of dimension 2xnxn.

What we seek to do here is to derive two different types of upper bounds. We will seek algorithms for ranking any tensor of a given dimension and will also attempt to find the maximum rank of any tensor of a given size and to demonstrate a tensor of that rank. We define the function r(m,p,q) as the maximum rank of any tensor of dimension mxpxq and establish the following properties of the function r(m;p,q).

Theorem 8: i) $r(m,p,q) = r(p,q,m) = r(p,m,q)$

ii) $m \leq r(m,p,q) \leq mp$

iii) $r(m,p,q_1+q_2) \leq r(m,p,q_1)+r(m,p,q_2)$

iv) $r(m,p,q_1q_2) \leq q_2r(m,p,q_1)$

Proof: i) follows from permutation equivalence and ii) follows from Theorem 3 of Chapter 2. iii) and iv) are obvious by considering the factorization (6). It is worthwhile to note that i) can be applied to ii), iii) and iv) to extend these results in a symmetric way.

Our plan of attack is to obtain methods of defining the function $r(m,p,q)$ which are also useful for ranking arbitrary tensors of dimension mxpxq.. We begin by studying the case $m = 2$ and seek methods of ranking the tensor (h_{ijk}) or of finding the degree of the characteristic function $G(s) = G_1s_1+G_2s_2$ where the jk element of G_i is (h_{ijk}) for $i=1,2$. The following results are useful for ranking most 2xnxn tensors.

Theorem 9: If $G(s) = G_1s_1+G_2s_2$ where G_1,G_2 are nxn matrices and if there exist scalars a,b,c,d such that aG_1+bG_2 and $\begin{bmatrix}ab\\cd\end{bmatrix}$ are nonsingular, then, if the Jordan canonical form for $(aG_1+bG_2)^{-1}(cG_1+dG_2)$ has p different 1-chains

$$\delta(G(s)) = n+p$$

Proof: From the results of Theorem 1 of Chapter 2, we know that if P and Q are invertible matrices, then $\delta(G(s)) = \delta(PG(Qs))$. The choice of $P = (aG_1+bG_2)^{-1}$, $Q = \begin{bmatrix}ab\\cd\end{bmatrix}$ and the change of variables to $t = Qs$ yields the result that $\delta(G(s)) = \delta(H(t))$ where $H(t) = H_1t_1+H_2t_2$, H_1 is the nxn identity matrix and $H_2 = (aG_1+bG_2)^{-1}(cG_1+dG_2)$. If we choose R and R^{-1} such that $RH_2R^{-1} = \Lambda_J$ is in Jordan canonical form, then $\delta(RH(t)R^{-1}) = \delta(H(t))$ and $RH(t)R^{-1} = I_nt_1+\Lambda_Jt_2$. We let $F(t) = RH(t)R^{-1}$ in what follows. Therefore, we can assume as a canonical form for this class of 2xnxn characteristic functions, the form

$$F(t)=\begin{bmatrix} \Delta_1(t) & & & & \\ & \Delta_2(t) & & & \\ & & \ddots & & \\ & & & \Delta_p(t) & \\ & & & & \Delta_{p+1}(t) \end{bmatrix}$$

$$\text{where } \Delta_i(t)=\begin{bmatrix} t_1+\lambda_i t_2 & t_2 & & \\ & t_1+\lambda_i t_2 & t_2 & \\ & & \ddots & t_2 \\ & & & t_1+\lambda_i t_2 \end{bmatrix}$$

$$\text{and } \Delta_{p+1}(t)=\begin{bmatrix} t_1+\lambda_{p+1}t_2 & & & \\ & t_1+\lambda_{p+2}t_2 & & \\ & & \ddots & \\ & & & t_1+\lambda_r t_2 \end{bmatrix}.$$

Because of the form of F(t) and $\Delta_i(t)$, we can extend the arguments of Section 1 of this chapter to prove that $\delta(F(t)) = \sum_{i=1}^{p+1} \delta(\Delta_i(t))$

and it is easily verified that if each Δ_i is an $n_i x n_i$ block then

$$\delta(\Delta_i(t)) = \begin{cases} n_i+1 \text{ if } 1\leq i\leq p \\ n_i \text{ if } i=p+1 \end{cases} . \text{ Therefore } \delta(F(t)) = n+p.$$

It is clear from this theorem that a wide class of tensors of dimension 2xnxn are of rank at most $\left\lfloor\frac{3n}{2}\right\rfloor$. We can extend this result to all tensors of this dimension.

<u>Theorem 10</u>: $r(2,n,n) = \left\lfloor\frac{3n}{2}\right\rfloor$

Proof: It is clear from Theorem 9 that $r(2,n,n) \geq \left\lfloor \frac{3n}{2} \right\rfloor$. We show that if $G(s)=s_1G_1+s_2G_2$ is an nxn characteristic function then there is an n-kxn-k characteristic function $H(s)$ of index 2 such that $\delta(G(s)) \leq \delta(H(s)) + \left\lfloor \frac{3k}{2} \right\rfloor$. If neither G_1 nor G_2 is of full rank, then we can assume $G(s) = \begin{bmatrix} I_p s_1 + As_2 & B_1s_2 & B_2s_2 \\ C_1s_2 & 0 & 0 \\ C_2s_2 & 0 & I_ts_2 \end{bmatrix}$ where A is in Jordan canonical form since this form can be reached from any G_1 and G_2 by factoring G_1 through the identity, transforming the upper corner of G_2 into Jordan canonical form and factoring the lower corner through the identity. It is clear that in this formulation $p > \left\lceil \frac{n}{2} \right\rceil$ or naive evaluation of G_1 and G_2 separately would yield the desired result.

For the canonical form given above, 2 cases arise depending on whether $t=0$ or $t > 0$. If $t > 0$, then we can write $G(s) = K_1(s)+K_2(s)$ where $K_1(s) = \begin{bmatrix} B_2C_2 & 0 & B_2 \\ 0 & 0 & 0 \\ C_2 & 0 & I_t \end{bmatrix} s_2$ and $K_2(s) = G(s)-K_1(s)$ is of dimension n-txn-t. By Theorem 8 of Chapter II, we know that $\delta(G(s)) \leq \delta(K_1(s))+\delta(K_2(s)) \leq \delta(K_2(s))+t$ and thus the theorem holds in this case. If $t = 0$, then either there is a value of i between 1 and n-p such that the ith row of B_1 and the ith column of C_1 contain nonzero elements or $\text{rank}(B_1)+\text{rank}(C_1) \leq n-p$ and if we denote the upper left corner of $G(s)$ by $H(s)$, then $\delta(G(s)) < \delta(H(s))+n-p$. We therefore assume that such an i exists and assume further that the ip elements of B_1 and C_1^T are nonzero. Letting $\underline{b}_i'$(resp. $\underline{c}_i$) denote the ith row(resp. column) of B_1(resp. C_1), we derive the decomposition $G(s) = M_1(s)+M_2(s)+M_3(s)+M_4(s)$ where

$$M_1(s) = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{i-1,i} \\ 1 \\ a_{i+1,i} \\ \vdots \\ a_{pi} \\ \underline{c}_i \end{bmatrix} [A_{i1} \cdots a_{i,i-1}, \; 1 \;\; a_{i,i+1} \cdots a_{i,p} \;\; \underline{b}_i']$$

$$M_2(s) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \leftarrow \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad [0 \;\ldots\; 0 \; \underset{\substack{\uparrow \\ \text{ith position}}}{1} \; 0 \;\ldots\; 0] \qquad (s_1 + (a_{ii} - 1) s_2)$$

$M_3(s) = -s_2 fg'$ where

$$F = \begin{bmatrix} b_{1,n-p} + a_{1i} (\underline{b}_i')_{n-p} \\ b_{2,n-p} + a_{2i} (\underline{b}_i')_{n-p} \\ \vdots \\ b_{i-1,n-p} + a_{i-1,i} (\underline{b}_i')_{n-p} \\ 0 \\ b_{p,n-p} + a_{pi} (\underline{b}_i')_{n-p} \\ \underline{c}_i' \end{bmatrix} \quad g = \frac{1}{(\underline{c}_i)_{n-p}} \begin{bmatrix} c_{n-p,1} + a_{i1} (\underline{c}_i')_{n-p} \\ \vdots \\ c_{n-p,i-1} + a_{i,i-1} (\underline{c}_i')_{n-p} \\ 0 \\ c_{n-p,p} + a_{i,p} \\ (\underline{b}_i)_1 \\ \vdots \\ (\underline{b}_i)_{n-p-1} \\ 1 \end{bmatrix}$$

then $M_4(s)$ is an n-2xn-2 characteristic function and

$$\delta(G(s)) \leqslant \sum_{i=1}^{4} \delta(M_i(s)) \leqslant 3+\delta(M_4(s))$$

By proceeding through the constructions given in this proof it is possible to find a realization of any nxn characteristic function of index 2 which is of degree not greater than $\left\lfloor \frac{3n}{2} \right\rfloor$ and often of smaller degree. Unfortunately, this construction does not always yield a realization of minimal degree and it is not possible to determine an analog of Theorem 9 which always yields a minimal degree realization. By extending these arguments to characteristic functions of index three it is possible to obtain the following results.

<u>Theorem 11</u>: i) $r(3,2,2) = 3$ and if $G(s)$ is any nondegenerate 2x2 characteristic function of index 3, then $\delta(G(s)) = 3$

ii) $r(3,3,3) = 5$

iii) $r(3,n,n) \leqslant 2n$ for all n

<u>Proof</u>: i) A special case of this result was known in 1932 to Oldenburger [18]. If $G(s) = G_1s_1+G_2s_2+G_3s_3$, then the result is trivial unless at least one of the G_i is nonsingular. Assume therefore that G_1 is nonsingular, then realizing $G(s)$ is equivalent to realizing $H(s)$ where $H(s) = I_2s_1+\Lambda s_2+Cs_3$ where Λ is in Jordan canonical form. Two separate cases occur depending on the form of Λ, if $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ and $C = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}$, then the factorization of degree 3 is

$$H(s) = \begin{bmatrix} 1 & 0 & c_2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} s_1+\lambda_1s_2+(c_1-c_2c_3)s_3 & & \\ & s_1+\lambda_2s_2+(c_4-1)s_3 & \\ & & s_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ c_3 & 1 \end{bmatrix}$$

If $\Lambda = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, then the degree 3 factorization is

$$\begin{bmatrix} 1 & 1 & 0 \\ \frac{c_3}{c_1-c_4} & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1+\lambda s_2+c_1 s_3 & & \\ & \lambda s_2+c_2 s_3 & \\ & & s_1+\lambda s_2+c_4 s_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{-c_3}{c_1-c_4} & 1 \end{bmatrix} \text{ if } c_1 \neq c_4$$

or

$$\frac{1}{2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} s_1+\lambda s_2+(c_1+c_3)s_3 & & \\ & s_1+\lambda s_2+(c_1-c_3)s_3 & \\ & & 2s_2+2(c_2-c_3)s_3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \text{if } c_1=c_4$$

The proofs of ii) and iii) involve detailed calculations and are included in the appendix.

Although the results presented in this section are interesting and insightful, it is clear that extension of these results would be a nontrivial exercise if the present methods were to be used. The complexity of these results suggests that new tools are necessary before an algorithm for ranking all tensors, or even a large subset of the set of all tensors, can be found. The results given here are useful for obtaining realizations of characteristic functions by decomposing intricate characteristic functions. Realizations generated in this manner are generally better than those which could be generated naively but are often non-minimal. This nonminimality arises because of intertwining of elementary characteristic functions when combined into intricate characteristic functions in such a manner as to yield a reduction in the degree of the latter. An understanding of this intertwining combined with our understanding of the elementary cases considered in this section may lead to better algorithms for determining characteristic function degrees and would in any case certainly provide improved heuristics. In the Appendix some elementary characteristic functions are studied in order to understand some of the heuristics which have been used in obtaining realizations.

CHAPTER IV

N-LINEAR EXTENSIONS

1. N-Linear Multiplication Problems

In this chapter, we consider the evaluation of a set of n-linear forms or the realization of an n-linear multiplication operation in terms of k-linear operations (k < n). We begin by generalizing the results of Chapter II in order to set the notation necessary to explore this problem. Although all the results given there can be generalized, we will only study those of independent interest in an n-linear setting. The operations which we consider are specified by a set of n-linear maps

$$h_i : \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \ldots \times \mathbb{R}^{p_n} \to \mathbb{R} \quad i=1,\ldots p_o \qquad (7)$$

Of the formulations given for bilinear operations in Chapter II, the most useful in this case are the defining function and tensor approaches. We let

$$f(\underline{x}_o, \underline{x}_1, \ldots \underline{x}_n) = \sum_{i=1}^{p_o} (\underline{x}_o)_i h_i(\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n)^{1} \qquad (8)$$

represent the defining function for the operation specified by (7) where each $\underline{x}_i$ is a p_i vector and $h_i(\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n)$ represents the value of the n-linear operation h_i applied to the vector $(\underline{x}_1, \ldots \underline{x}_n)$. The defining function gives rise to a tensor E of order n+1 where we define the

[1] In this chapter, each vector is underlined and $(\underline{x}_o)_i$ represents the ith component of the vector $\underline{x}_o$.

$k_o,k_1,\dots k_n$ component of E by the relation

$$(E)_{k_o k_1 \dots k_n} = \frac{\partial^{n+1} f}{\partial(\underline{x}_o)_{k_o} \partial(\underline{x}_1)_{k_1} \dots \partial(\underline{x}_n)_{k_n}} \tag{9}$$

It is worthwhile to note that the correspondence of the tensor E to the operation (7) could also have been established by using the relation

$$h_i(\underline{x}_1,\underline{x}_2,\dots \underline{x}_n) = \sum_{k_1=1}^{p_1} \sum_{k_2=1}^{p_2} \dots \sum_{k_n=1}^{p_n} E_{ik_1 \dots k_n} \prod_{j=1}^{n} (\underline{x}_j)_{k_j} \tag{10}$$

Throughout this chapter we shall use all three of these formulations interchangably as it is clear that equations (7) - (10) can be used to relate the various formulations.

We shall begin our study of n-linear operations by considering realizations which are extensions of the realizations which we sought for bilinear operations. Thus, we shall seek factorizations of the form

$$f(\underline{x}_o,\dots \underline{x}_n) = \sum_{i=1}^{m} \prod_{j=0}^{n} \langle \underline{a}_{ij},\underline{x}_j \rangle \tag{11}$$

or

$$E_{k_o,k_1,\dots k_n} = \sum_{i=1}^{m} \prod_{j=0}^{n} (\underline{a}_{ij})_{k_j} \tag{12}$$

where the $\underline{a}_{ij}$ are vectors belonging to a subset of K of $\mathbb{C}$. We shall define the degree of a set of n-linear forms (7) over $K \subset \mathbb{C}$ as the least m such that a factorization of the form (11) (or equivalently (12)) exists with all elements of $\underline{a}_{ij}$ belonging to K. We will denote the degree by $\delta_{K,n}(h_1,\dots h_{p_o})$ or $\delta_{K,n}(f)$ or $\delta_{K,n}(E)$. It is clear that $\delta_{K,n}(E)$ is equal to the rank of the tensor E over K. In the case where n=2 we shall write δ_K for $\delta_{K,2}$. The n+1-tuple $(p_o,p_1,\ldots p_n)$ will represent the dimension of the operation described by (7) - (10). This quantity is often referred to as the valence set of the tensor E.

With this formulation, it is possible to present the notion of permutation equivalence in its most general form. We will refer to the tensor defined by (9) as the standard tensor corresponding to the operation of defining function (8). If π is any element of S_{n+1}, the permutation group on n+1 elements, and E is a tensor of dimension $(p_o,p_1,\ldots p_n)$, we will let $\pi(E)$ represent the tensor of dimension $(p_{\pi(0)},p_{\pi(1)},\ldots p_{\pi(n)})$ such that the $i_o,i_1,\ldots i_n$ element of E is the $i_{\pi(0)},i_{\pi(1)},\ldots i_{\pi(n)}$ element of $\pi(E)$. We can now state our permutation equivalence result for this class of realizations.

<u>Theorem 1</u>: If E is the standard tensor of order n+1 corresponding to an n-linear operation of defining function f, and $\pi \in S_{n+1}$, then for all $K \subset \mathbb{C}$,

$$\delta_{K,n}(E) = \delta_{K,n}(\pi(E))$$

<u>Proof</u>: Observe that E is generated from the defining function f via (9) and $\pi(E)$ is generated by letting the $i_o,i_1,\ldots i_n$ element of $\pi(E)$ be given by $\dfrac{\partial^{n+1} f}{(\partial \underline{x}_{\pi(0)})_{i_o}(\partial \underline{x}_{\pi(1)})_{i_1}\cdots(\partial \underline{x}_{\pi(n)})_{i_n}}$. A factorization of the form (11) for f yields a factorization of the form (12) for E and a factorization of the form $(\pi(E))_{k_o,k_1,\ldots k_n} = \sum_{i=1}^{m}\sum_{j=0}^{n} a_{ik_{\pi(j)}}$ for $\pi(E)$. A total of (n+1)! n-linear operations are related in this way.

Following the notation for bilinear operations, we shall denote by $M_{\nu_1,\nu_2,\ldots\nu_n,\nu_{n+1}}$, the tensor associated with the n-linear operation of computing the product of $\nu_1 x \nu_2$, $\nu_2 x \nu_3,\ldots \nu_{n-1} x \nu_n$ and $\nu_n x \nu_{n+1}$ matrices and by $M_{\nu_1,\ldots\nu_{n+1}}(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ the corresponding defining function, where $\underline{x}_i$ is a $\nu_i\nu_{i+1}$ vector for i=1,...n and $\underline{x}_o$ is a $\nu_1\nu_{n+1}$ vector. $P_{\mu_1,\mu_2,\ldots\mu_n}$ will represent the tensor associated with the n-linear operation of computing the product of polynomials of degrees μ_1-1, $\mu_2-1,\ldots\mu_n-1$ and $p_{\mu_1,\ldots\mu_n}(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ the corresponding defining function where $\underline{x}_i$ is a μ_i vector for i=1,...n and $\underline{x}_o$ is a $\mu_1+\mu_2+\ldots+\mu_n-(n-1)$ vector. These operations have similar structure and properties in an n-linear setting as bilinear versions had as will be discussed below.

Of particular interest in an n-linear setting are the determinant and permanent operations as n-linear functions of the row or column vectors of an nxn matrix. We will denote the tensors corresponding to these operations by D_n and Q_n respectively and the defining functions by $d_n(\underline{x}_1,\ldots\underline{x}_n)$ and $q_n(\underline{x}_1,\ldots\underline{x}_n)$ where $\underline{x}_i$ is an n-vector for i=1,...n. Since, the determinant (or permanent) of a matrix is an n-linear operation yielding only one output, it is (permutation) equivalent to an n-1-linear operation yielding n outputs. Before proceeding, it is of value to identify the structure of D_n and Q_n and we observe that the $i_1,i_2,\ldots i_n$ element of D_n is zero unless $i_1\ldots i_n$ are all different in which case $(D_n)_{i_1\ldots i_n}$ is +1 if $i_1\ldots i_n$ is obtained from 1...n by an even permutation and -1 otherwise. The $i_1\ldots i_n$ element of Q_n is equal to the absolute value of the corresponding element of D_n. We can now characterize interesting operations which are permutation equivalent to D_n and Q_n.

Theorem 2: i) Computing the determinant of an nxn matrix by an algorithm of the form studied here is equivalent to computing the determinants of all n-1xn-1 submatrices of an n-1xn matrix by an algorithm of the form studied here.

ii) Computing the permanent of an nxn matrix by an algorithm of the form studied here is equivalent to computing the permanents of all n-1xn-1 submatrices of an n-1xn matrix by an algorithm of the form studied here.

Proof: Let $d_n(\underline{x}_1,\underline{x}_2,\ldots\underline{x}_n)$ (resp. $q_n(\underline{x}_1,\underline{x}_2,\ldots\underline{x}_n)$) be the defining function corresponding to D_n(resp. Q_n). Then

$$d_n(\underline{x}_1,\underline{x}_2,\ldots\underline{x}_n) = \sum_{\pi\epsilon S_n} \mathrm{sgn}(\pi) \prod_{i=1}^{n} (\underline{x}_i)_{\pi(i)} = \sum_{j=1}^{n} (\underline{x}_1)_j \sum_{\substack{\pi\epsilon S_n \\ \pi(1)=j}} \mathrm{sgn}(\pi) \prod_{k=2}^{n} (\underline{x}_k)_{\pi(k)}$$

corresponding to computing the determinants of all n-1xn-1 submatrices of the matrix with rows $\underline{x}_2,\underline{x}_3,\ldots\underline{x}_n$. The same argument holds for Q_n and $q_n(\underline{x}_1,\underline{x}_2,\ldots\underline{x}_n)$ with the $\mathrm{sgn}(\pi)$ term eliminated from all sums.

This theorem is interesting in two respects, it provides a first insight into a method of determinant and permanent calculations (i.e. expansion by minors) and it presents a view of similarities between the operations. Both of these points will be discussed at length below.

2. Equivalence, Dominance and Reduction

As was shown in Chapters II and III, it is of value to have methods for defining a partial ordering on the set of tensors of a given order with respect to $\delta_{K,n}$. In the last section, we extended the notion of permutation equivalence and we begin here by extending the dominance results given for bilinear operations. If there exist matrices $P_o,P_1,\dots P_n$ over K such that $f(P_o\underline{y}_o,P_1\underline{y}_1,\dots P_n\underline{y}_n) = g(\underline{y}_o,\underline{y}_1,\dots\underline{y}_n)$ then we say that defining function f structurally dominates defining function g over K (i.e. $f \supsetneq_K g$). This amounts to being able to code computations of g into computations of f. If the coding is bijective (i.e. $f \supsetneq_K g$, $g \supsetneq_K f$) corresponding to the inverse matrices $P_o^{-1},P_1^{-1},\dots P_n^{-1}$ existing over K, then we say that f and g are structurally equivalent and write $f \sim_K g$. It is easy to translate these results into tensor notation and we shall use the notation $F \supsetneq_K G$ (resp. $F \sim_K G$) if F and G are tensors corresponding to defining functions f and g such that $f \supsetneq_K g$ (resp. $f \sim_K g$). It is easy to establish the following extension of the first theorem of Chapter II.

Theorem 3: Let K be a subring of $\mathbb{C}$, then

i) If $f \supsetneq_K g$ then $\delta_{K,n}(f) \geq \delta_{K,n}(g)$

ii) If $f \sim_K g$ then $\delta_{K,n}(f) = \delta_{K,n}(g)$

Proof: If $\delta_{K,n}(f) = m$, then the minimal factorization of $f(\underline{x}_o,\dots\underline{x}_1)$ of the form (11) over K is of degree m. Then, if $f(P_o\underline{y}_o,P_1\underline{y}_1,\dots,P_n\underline{y}_n) = g(\underline{y}_o,\underline{y}_1,\dots\underline{y}_n)$, there is a factorization of g of degree m, given by

$$g(\underline{y}_o,\underline{y}_1,\dots\underline{y}_n) = \sum_{i=1}^{m}\prod_{j=0}^{n}\langle\underline{a}_{ij},P_j\underline{y}_j\rangle = \sum_{i=1}^{m}\prod_{j=0}^{n}\langle P_j^T\underline{a}_{ij},\underline{y}_j\rangle$$

The theorem follows from this factorization.

When considering tensors of different orders a new type of dominance arises. We will say that a defining function $f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ tensorially dominates a defining function $g(\underline{y}_o,\underline{y}_1,\ldots\underline{y}_p)$ $(f \overset{T}{\supset} g)$ if $p=n-1$ and $f(\underline{x}_o,\underline{x}_1,\ldots,\underline{x}_{n-1},\underline{x}_n) = g(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_{n-2},\underline{x}_{n-1} \otimes \underline{x}_n)$ or if there exists a defining function h such that $f \overset{T}{\supset} h$ and $h \overset{T}{\supset} g$. Intuitively, if $f \overset{T}{\supset} g$, then an algorithm for evaluating g is a method for evaluating f by forming sums of linear and multilinear forms in $\underline{x}_o,\ldots\underline{x}_p$. In assessing the value of algorithms of this sort, two considerations are necessary. The first regarding the relative degrees of f and g if $f \overset{T}{\supset} g$ is studied here, and the second regarding the complexity of algorithms for computing f and g if $f \overset{T}{\supset} g$ is discussed in a later section. As might be expected, tensorial dominance implies degree dominance but an upper bound on this dominance exists.

<u>Theorem 4</u>: Let $f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ be a defining function corresponding to a tensor of dimension $(p_o,p_1,\ldots,p_{n-1},p_n)$ and $g(\underline{y}_o,\underline{y}_1,\ldots,\underline{y}_{n-1})$ be a defining function corresponding to a tensor of dimension $(p_o,p_1,\ldots p_{n-2},p_{n-1}p_n)$ such that $f \overset{T}{\supset} g$. Then, for all $K \subset \mathbb{C}$,

$$\delta_{K,n-1}(g) \leqslant \delta_{K,n}(f) \leqslant \min(p_n,p_{n-1})\delta_{K,n-1}(g)$$

<u>Proof</u>: If $f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ is of degree m and the factorization of (11) holds, then by using the identity $<\underline{a}_{i,n-1}\underline{x}_{n-1}><\underline{a}_{in},\underline{x}_n> = <\underline{a}_{i,n-1}\otimes\underline{a}_{in},\underline{x}_{n-1}\otimes\underline{x}_n>$ a degree m factorization of $g(x_o,x_1,\ldots x_{n-1}\otimes\underline{x}_n)$ is found.

If $g(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_{n-1} \otimes \underline{x}_n) = \sum_{i=1}^{r} (\prod_{j=0}^{n-2} <\underline{a}_{ij},\underline{x}_j>)(<\underline{a}_{i,n-1},\underline{x}_{n-1} \otimes \underline{x}_n>)$ then each term of the form $<\underline{a}_{i,n-1},\underline{x}_{n-1} \otimes \underline{x}_n>$ can be written as $\sum_{i=1}^{q} <\underline{b}_i,\underline{x}_{n-1}><\underline{c}_i,\underline{x}_n>$ where $q \leqslant \min(p_n,p_{n-1})$ yielding a factorization of $f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ of degree at most $\min(p_n,p_{n-1})\delta_{K,n-1}(g)$.

This theorem and Theorem 3 of Chapter I yield

<u>Corollary</u>: If E is a tensor of dimension $(p_o, p_1, \ldots p_n)$ over K and $p = \max(p_o, p_1, \ldots p_n)$, then

$$p \leq \delta_{K,n}(E) \leq \frac{\prod_{i=0}^{n} p_i}{p}$$

We can use these results to prove the following results about n-linear operations.

<u>Theorem 5</u>: The following are true for all K

i) $\delta_{K,n}(D_n) \geq (n-1)^{\lfloor n/2 \rfloor}$

ii) $\delta_{K,n}(Q_n) \geq (n-1)^{\lfloor n/2 \rfloor}$

iii) $\delta_{K,n}(P_{d_1,\ldots d_n}) \geq \delta_{K,n-1}(P_{d_1+d_2-1,d_3,\ldots d_n}) \geq \ldots \geq \delta_{K,2}(P_{d_1+\ldots+d_{n-1}-(n-2),d_n})$

iv) For all $\pi \in S_{n+1}$,

$$\delta_{K,n}(M_{d_o,d_1,\ldots d_n}) = \delta_{K,n}(M_{d_{\pi(0)},d_{\pi(1)},\ldots d_{\pi(n)}})$$

<u>Proof:</u> i) and ii) result from reducing D_n and Q_n from nth order tensors of dimension $(n,n,n,\ldots,n)$ to second order tensors (i.e. matrices) of dimension $(n^{\lfloor n/2 \rfloor}, n^{\lceil n/2 \rceil})$ and observing that these matrices are of rank $(n-1)^{\lfloor n/2 \rfloor}$.

iii) is true because there is a third order tensor H_{d_1,d_2,d_3} such that $P_{d_1,d_2,d_3} \overset{T}{\hookrightarrow} H_{d_1,d_2,d_3} \underset{K}{\hookrightarrow} P_{d_1+d_2-1,d_3}$ and iv) is true since any element of S_{n+1} can be written as the product of transpositions and cycles of length 3 and by the result of Hopcroft and Musinski [12] which this generalizes, $\delta_{K,n}(M_{d_o,d_1,\ldots d_n})$ is invariant under transpositions and cycles of length 3.

Corollary: If K is any subfield of $\mathbb{C}$, then

$$\delta_{K,n}(P_{d_1,d_2,\ldots d_n}) = d_1+d_2+\ldots+d_n-(n-1)$$

Proof: By iii) of the theorem, $\delta_{K,n}(P_{d_1,d_2,\ldots d_n}) > \delta_K(P_{d_1+d_2+\ldots+d_{n-1}-(n-2),d_n})$ and therefore, $\delta_{K,n}(P_{d_1,d_2,\ldots d_n}) \geq d_1+d_2+\ldots+d_n-(n-1)$. Equality comes by letting $\alpha_1,\ldots\alpha_d$ be distinct elements of K for $d = \sum_{i=1}^{n}(d_i-1)+1$. Then, if $(\underline{a}_{ik_j})_s = \alpha_i^{s-1}$ for $1 \leq i \leq d$, $1 < s < d_i$, $1 < j < n$, a choice of $\underline{a}_{ik_o}$ exists for $1 \leq i \leq d$ such that a factorization of the form (12) holds for $P_{d_1,d_2,\ldots d_n}$. The vector $\underline{a}_{ik_o}$ is the ith row of the inverse of the dxd matrix with ijth element α_i^{j-1}.

The first result of Theorem 5 points out a limitation of the model we have chosen to realize n-linear multiplication operations. The well known Gaussian elimination algorithm computes the determinant of an arbitrary nxn matrix in $O(n^3)$ arithmetic operations which is much less than the number given here of at least $O(\sqrt{n!})$. Therefore, it is necessary to examine what part of the structure of the model proposed here prevents this bound from being reached. The remainder of this chapter is devoted to outlining methods of studying possible alternatives to this model which are more flexible than the one given here. We study two models that do not have the restriction that all intermediate products must involve n terms.

It is worthwhile to note that in other cases the present model seems to yield reasonable algorithms and to consider the structures of the determinant and permanent operations. Knuth [15, p.426] notes that while Gaussian elimination computes determinants in $O(n^3)$ operations, no algorithm is known for computing the permanent of a matrix which grows at slower than an exponential rate, and thus the result we obtained of $O(n^{n/2})$ as a lower bound on permanent calculations via n-linear systems may also be valid for all other models of computation. We extend a problem given by Knuth [15] to the following problem

Open Problem: Given $T_n \subset S_n$, a family of subsets of order growing asymptotically with $k(n)$, and g an integer valued function defined on S_n. Let $c(g,T_n)$ represent the complexity of computing $\sum_{\pi\varepsilon T} g(\pi) \prod_{i=1}^{n} a_{i\pi(i)}$ for arbitrary nxn matrices $A = (a_{ij})$. For what sets T_n and functions g, is the growth rate of $c(g,T_n)$ slower than $O(k(n))$?

At present, the only known set and function for which such an algorithm is known is the determinant calculation for which $T_n = S_n$ and $g(\pi) = \text{sgn}(\pi)$ which is +1 if π is an even permutation and -1 if π is odd. It is clear that a "faster" algorithm exists for T_n the set of even (or odd) permutations in S_n and $g(\pi) = 1$ for all π, if and only if one exists also for permanent calculations.

3. Alternate Formulation I - Interconnections of k-linear Systems

In the last section we considered schemes of the form in Figure 1a) for realizing n-linear multiplication operations with inputs $\underline{x}_1,\underline{x}_2,\ldots\underline{x}_n$ and showed that while such systems had nice properties, they were not optimal for all applications. We consider here systems of the form given in Figure 1b) as an extension of these systems. In considering such systems, it is necessary to slightly expand the notation introduced above. If A_r is an $m\mathrm{x}p_r$ matrix ith row given by $\underline{a}_{ir}$ for r=0,1,...n and a factorization of the form (11) (or (12)) holds, then we will say that $(A_o,A_1,\ldots A_n)$ forms an n-linear multiplication system (nms) for f(or E or $\{h_i\}$). It is easily verified that the factorization in (12) can be rewritten as

$$E_{k_o,\ldots,k_n} = \sum_{i=1}^{m}\prod_{j=0}^{n}(A_j)_{ik_j} \tag{12'}$$

What we seek to do below is to outline methods of characterizing the analog of (12') for realizations of the form of Figure Ib).

We will say that the scheme of Figure Ib) represents the interconnection of $(k_1,\ldots,k_r)$ linear systems $(I, B_{11},\ldots B_{1k_1})$, $(I, B_{21},\ldots B_{2k_2}),\ldots (I, B_{r1},\ldots B_{rk_r})$ through the r-linear system $(C_o,C_1,C_2,\ldots C_r)$. If the systems of Figures Ia) and b) realize the same n-linear operation then

$$\sum_{i=0}^{m}\prod_{j=0}^{n}(A_j)_{i\ell\Lambda} \quad \sum_{i_1=1}^{\delta_1}\cdots\sum_{i_r=1}^{\delta_r}\Gamma_{\ell_o,i_1,\ldots i_r}\prod_{\alpha=1}^{r}\prod_{\beta=1}^{k_\alpha}(B_{\alpha\beta})_{i,\ell_{\sum_{\gamma=0}^{r-1}k_\gamma+\beta}} \tag{13}$$

where

$$\Gamma_{\ell_o,i_1,\ldots i_r} = \sum_{i=1}^{\delta_o}(C_o)_{\ell_o i}\prod_{j=1}^{r}(C_j)_{ii_j}$$

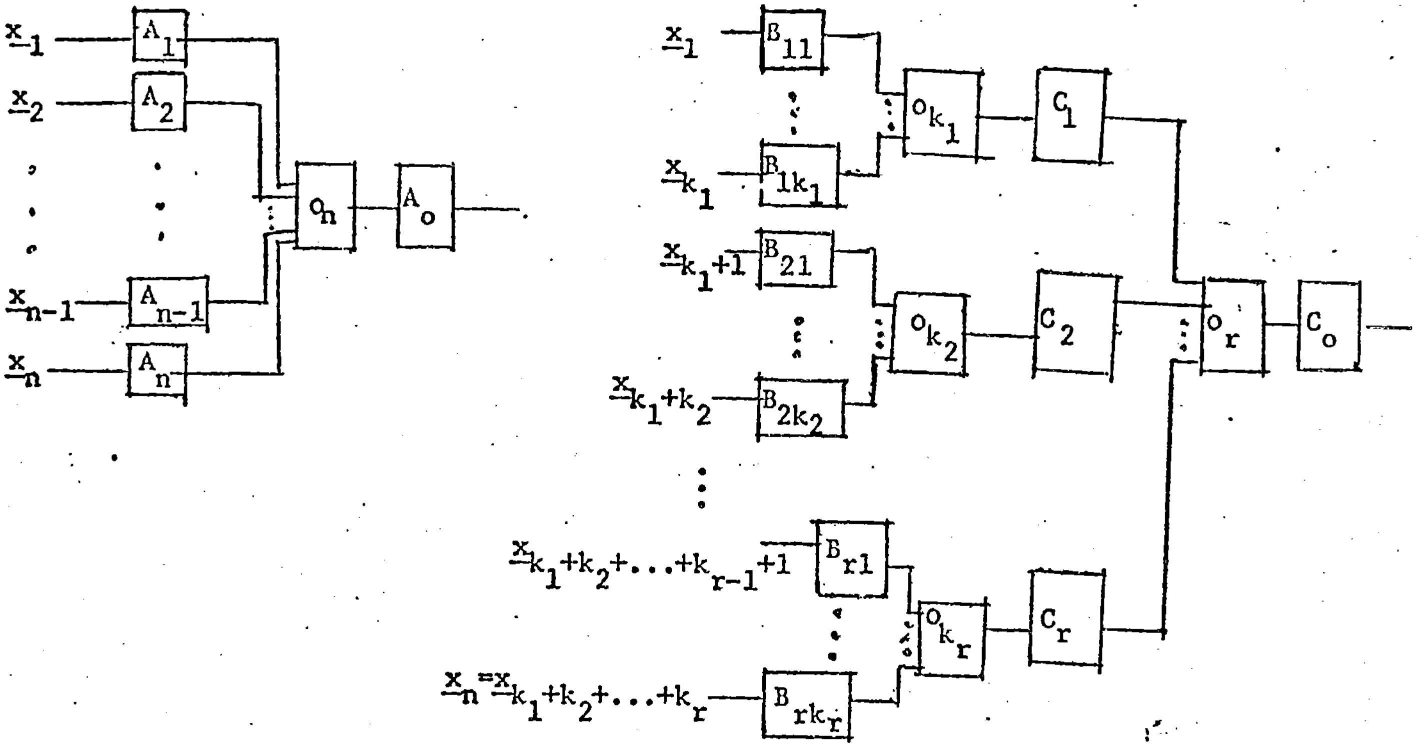


Figure I: The 2 linear models for realizing n-linear multiplication operations. Blocks of the form $\underline{x}_i - \boxed{D} -$ compute $D\underline{x}_i$, a vector of linear forms in $\underline{x}_i$ and blocks of the form $\boxed{o_p}$ compute pointwise (Haddamard) products of a set of pinput vectors.

a) $(A_o, A_1, \ldots A_n)$ is an nms operating on inputs $\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n$

b) $(I, B_{i1}, \ldots B_{ik_i})$ is a k_i-ms operating on inputs $\underline{x}_{k_1+\ldots+k_{i-1}+1}, \ldots \underline{x}_{k_1+\ldots+k_i}$ and $(C_o, C_1, \ldots C_r)$ is an rms working on the outputs of the r interconnected systems.

The systems $(I, B_{j1},\ldots B_{jk_j})$ are of dimension δ_j and the system $(C_o,C_1,\ldots C_r)$ is of dimension δ_o. In this case we say that the defining function f realized by (12') is also realized by a δ_o-dimensional connection of k_i-linear systems of dimensions δ_i. We will let the vector $(\delta_1,\ldots\delta_r,\delta_o)$ represent the dimension vector of this realization and will define the K-cost of the realization by the 2-tuple (γ_K^1,γ_K^2) where $\gamma_K^1 = \delta_o+\delta_1+\delta_2+\ldots+\delta_r$ and $\gamma_K^2 = \delta_1(k_1-1)+\delta_2(k_2-1)+\ldots+\delta_r(k_r-1)+\delta_o(r-1)$. The first of these numbers gives a measure of the size of the realization and the second a measure of its complexity. If $r=1$ these numbers reveal the same quantity, however for larger values of r, a tradeoff similar to time-storage tradeoffs introduced elsewhere (e.g. [19]) is possible. We shall define by $\beta^1_{K,(k_1,\ldots k_r),r}(f)$ (resp. $\beta^2_{K,(k_1,\ldots k_r),r}(f)$) the minimum value of γ_K^1 (resp. γ_K^2) for all realizations of f formed by the interconnections of $(k_1,k_2,\ldots,k_r)$-linear systems and by $\beta_K^1(f)$ (resp. $\beta_K^2(f)$) the minimum value of $\beta^1_{K_r(k_1,\ldots k_r),r}(f)$ (resp. $\beta^2_{K,(k_1,\ldots,k_r),r}(f)$) over all partitions of n into $\sum_{i=1}^{r} k_i$ where $k_i \geq 2$ for all i. The interesting tradeoff between values of $\beta_K^1(f)$ and $\beta_K^2(f)$ corresponding to a tradeoff between size and complexity is considered below.

<u>Lemma</u>: Let $(\delta_o,\delta_1,\ldots\delta_r)$ be the dimension of a realization of f formed by the interconnections of $(k_1,\ldots k_r)$-linear systems through an r-linear system which is minimal in that no realization of dimension $(\delta_o',\delta_1',\ldots\delta_r')$ exists for which at least one of $\delta_o-\delta_o',\delta_1-\delta_1',\ldots\delta_r-\delta_r'$ is positive and the rest are nonnegative. Then,

$$\max(\delta_o,\delta_1,\ldots\delta_r) \leq \beta^1_{K,(n),1}(f) = \delta_{K,n}(f) \leq \prod_{i=0}^{r} \delta_i$$

Proof: By examining (13) and (12') it is obvious that a factorization of the form in (13) gives rise to one of dimension $\delta_o \prod_{i=1}^{r} \delta_i$ in (12') and a factorization of the form in (12') gives rise to a factorization of the tensor of each system interconnected in (13) (or Figure Ib)).

Theorem 6: $\beta_K^2(f) \leqslant \beta_{K,(n),1}^2(f) = (n-1)\delta_{K,n}(f)$ for all K, and f(n > 2).

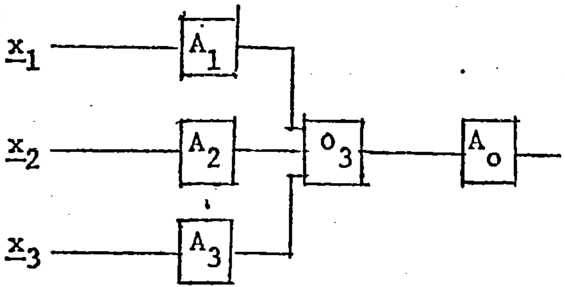
Proof: Define $\Delta_{k_1,\ldots k_r}(f) = \beta_{K,(k_1,\ldots k_r),r}^2(f) - \beta_{K,(n),1}^2(f)$ and we show that $\Delta_{k_1,\ldots k_r}(f) \leqslant 0$ for all $k_1,\ldots k_r$. Recall that $\beta_{K,(k_1,\ldots k_r),r}^2$ can be expressed as $\delta_o(r-1) + \sum_{i=1}^{r} \delta_i(k_i-1)$ where $\sum_{i=1}^{r} k_i = \sum_{i=1}^{r} (k_i-1)+r=n$ and $\max(\delta_o,\delta_1,\ldots\delta_r) \leqslant \delta_{K,n}(f)$ and $\beta_{K,(n),1}^{(2)}(f) = (n-1)\delta_{K,n}(f)$. Therefore $\Delta_{k_1,\ldots k_r}(f)$ can be rewritten as $(\delta_o - \delta_{K,n}(f))(r-1) + \sum_{i=1}^{r} (\delta_i - \delta_{K,n}(f))(k_i-1)$ a sum of non-positive terms.

The result of this theorem shows that the least complex realization is always a realization consisting of interconnections of bilinear systems. Further study is necessary to develop methods of finding such realizations, but it seems that an algorithm for computing determinants faster than $O(\sqrt{n!}\,)$ exists within this framework. This further study should be parallel to the extensions suggested at the end of the last chapter. For both of these projects it is necessary to gain a thorough understanding of basic bilinear characteristic functions (3rd order tensors), there it was necessary to study methods of interconnecting such functions to generate larger bilinear characteristic functions, and here we seek to find methods of interconnection to generate n-linear operations. However, it is further necessary to interconnect bilinear systems to realize n-linear system in such a way as to not cause drastic increases in the size of $\beta_K^1(f)$.

We close with an example of the two types of n-linear systems we have discussed.

Example: Let $X = \begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \underline{x}_3' \end{bmatrix} = (X_{ij})$ be a 3x3 matrix and suppose we wish to compute det(X), then by the methods of the previous section, we can

define the system

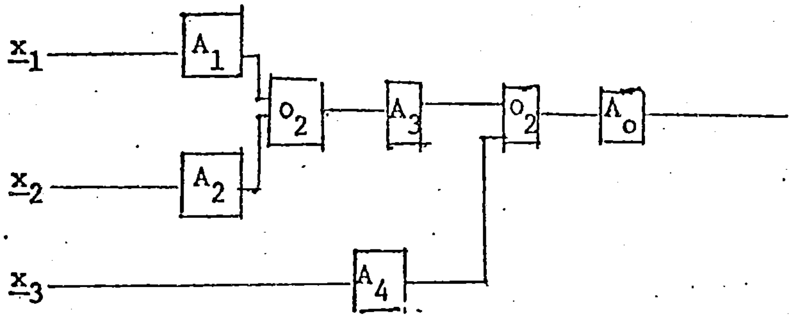


to form det(X)

and show that the minimal system has degree 5 and is given by

$$A_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & -1 \\ -1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_o = [1\ 1\ 1\ 1\ 1]$$

This system required 10 multiplications. An alternative system of the type discussed in the present section is given by



where

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}, \quad A_4 = I_3, \quad A_o = [1\ 1\ 1]$$

and this system as the interconnection of a degree 6 system with a degree 3 system requires 9 multiplications.

4. Alternate Formulation II-Inhomogeneous Systems

Another approach to the realization of n-linear operations by k-linear operations is to consider inhomogeneous systems. These systems are also of interest in the study of linear operations of different degrees. The value of an inhomogeneous system is that it allows the computation of k-linear forms through a model which appears to be n-linear and the computation of different set of k-linear forms in overlapping but non-identical inputs. This has the benefit of yielding a tractable computing model capable of doing more computations than the one discussed in the previous section.

Let $f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)$ be the defining function for an n-linear operation and let $\underline{x}_i$ be a p_i-vector for i=1,2,...n. We define the p_i+1-vectors $\hat{\underline{x}}_i = \begin{bmatrix}\underline{x}_i\\1\end{bmatrix}$ and say that $(A_o,A_1,\ldots A_n)$ defines an n-linear inhomogeneous multiplication system (nims) of dimension m, if A_o is an mxp_o matrix and each A_i is an $mx(p_i+1)$ matrix such that

$$f(\underline{x}_o,\underline{x}_1,\ldots\underline{x}_n)=f(\underline{x}_o,\hat{\underline{x}}_1,\hat{\underline{x}}_2,\ldots\hat{\underline{x}}_n) = \sum_{i=1}^{m} <\underline{a}_{io},\underline{x}_o> \prod_{j=0}^{n} <\underline{a}_{ij},\hat{\underline{x}}_j>$$

where $\underline{a}'_{ij}$ represents the ith row of the matrix A_j.

A cascade of n-ims's has the advantage of allowing the computation of interconnections of k_i-linear systems such that different k_i-linear systems operate on the same input vectors. This was not possible in the model proposed in the last section. If we extend this model further to consider systems which accept n input vectors and output n vectors, it is possible to simulate a form of Gaussian elimination. We can construct a system which accepts the row vectors of an nxn matrix and outputs after $O(n^3)$ multiplication, the numbers Δ_1 and Δ_2 such that the determinant of the matrix is Δ_1/Δ_2. However, no such cascade is known at present for permanent calculations.

This model is also of value for studying operations which have outputs which are n_i-linear functions of the inputs, but for which n_i varies over the outputs. An example of such an operation is the computation of the symmetric functions for a set of n numbers $\{x_1,\ldots x_n\}$ which was studied by Strassen [23]. The goal here is to find a method of computing the n-vector of outputs given by

$$\{\sum_{i=1}^{n} x_i,\ \sum_{i_1<i_2} x_{i_1}x_{i_2},\ldots,\ \sum_{i_1<i_2<..<i_p} x_{i_1}x_{i_2}\ldots x_{i_p},\ldots,\ \prod_{i=1}^{n} x_i\}$$

The first element of this vector is a linear function of the input, the second bilinear,... the ith i-linear, and the nth n-linear. With "mixed linearity" of this type, if we consider the input vectors as the n 2-vectors $\hat{\underline{x}}_1 = \begin{bmatrix} x_1 \\ 1 \end{bmatrix},\ldots\ \hat{\underline{x}}_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix},\ ..\ \hat{\underline{x}}_n = \begin{bmatrix} x_n \\ 1 \end{bmatrix}$, the output vector is an n-vector which is an n-linear function of $\hat{\underline{x}}_1,\ldots\hat{x}_n$ and we can find a tensor of order n+1 and dimension $(n,2,2,\ldots 2)$ which defines this operation. The $i_o,i_1,\ldots,i_n$ element of this tensor is one if $i_o = \sum_{j=1}^{n}(2-i_j)$ and zero otherwise. By a cascade of nims's it is possible to simulate the $O(n\log^2 n)$ algorithm given by Borodin [1] to perform this operation. The method presented here is inherently the same in this case as computing the product of the polynomials $(x-x_1),(x-x_2),\ldots)x-x_i),..$ and $(x-x_n)$, but in cases where embedding into a polynomial product (or other totally n-linear operation) is not obvious, the inhomogeneous model is of value.

CHAPTER V

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The research presented in this dissertation was motivated by the need for a setting in which previous results on the optimal evaluation of bilinear operations could be studied. Such a setting has been provided and used to interpret previous results in a new light. It has been possible to generalize many previous results in order to obtain new results as well as to introduce new methods for studying bilinear operations. Although many of the new results presented here are of independent importance, we feel that the major value of this dissertation is the model proposed. At the heart of this model is the introduction of the indeterminates s_i yielding a three-dimensional model into which both inputs and the output enter to replace previous two-dimensional models which only considered the inputs. The properties of such a model are studied and many important concepts are introduced for use in the study of the arithmetic complexity of bilinear operations. These ideas are helpful in understanding the complexity of a bilinear operation by determining in a more efficient manner than previously possible upper and lower complexity bounds. Methods of generating classes of realizations of an operation over various subsets of the complex field are discussed and a partial ordering on operations is defined. Tables of best known upper and lower bounds on some important bilinear operations are given in the appendix, although the main thrust of this dissertation has been to determine methods of improving all of these numbers rather than any individual result.

As is the nature of scientific research, this dissertation raises more questions than it settles. Along these lines, many open questions have been proposed within the body of the dissertation. Among the further avenues along which these results may be extended are:

1. An algebraic approach consisting of determining the degree of each of a set of elementary characteristic functions and then methods of determining degrees of characteristic functions consisting of interconnections of elements of this set as a function of the elementary function degrees and the types of interconnections.

2. A combinatorial approach which might consist of understanding how similarities and differences in the structures of the G_i affect the value of the degree of $G(s) = \Sigma G_i s_i$.

3. A framework in which upper and lower bounds can be obtained on operations applied to inputs such that linear forms in the inputs commute.

4. A study of methods of changing the basic operations from + and x to + and max in order to find a similar framework for sorting algorithms and a variety of combinational problems. Applications of this approach could be used to relate sorting a set of n numbers to determining the elementary symmetric functions for a set of n indeterminates. It may also be possible to relate matrix permanent calculations to the marriage problem.

REFERENCES

[1] A. Borodin, "On the Number of Multiplications Necessary to Compute Certain Functions, circa May, 1973", Proceedings of the CMU Symposium on the Complexity of Sequential and Parallel Numerical Algorithms, May, 1973.

[2] A. Borodin and R. Moenck, "Fast Modular Transformations", to appear.

[3] R. Brockett and D. Dobkin, "On the Optimal Evaluation of a Set of Bilinear Forms", Proceedings of the Fifth Annual ACM Symposium on the Theory of Computing, Austin, Texas, 1973, pp. 88-95.

[4] L. Calabi and E. Myrvaagnes, "On the Minimal Weight of Binary Group Codes", IEEE PGIT, Vol. 10, pp. 385-7, 1964.

[5] C.M. Fiduccia, "Fast Matrix Multiplication", Proceedings of the Third Annual ACM Symposium on the Theory of Computing, Shaker Heights, Ohio 1971, pp. 45-49.

[6] C.M. Fiduccia, "On Obtaining Upper Bounds on the Complexity of Matrix Multiplication", in Complexity of Computer Computations, (R. Miller and J. Thatcher, editors), Plenum Press, 1972.

[7] M.J. Fischer and L.J. Stockmeyer, "Fast On-Line Integer Multiplication", Proceedings of the Fifth Annual ACM Symposium on the Theory of Computing, Austin, Texas, 1973, pp. 67-72.

[8] F. Gantmacher, Applications of the Theory of Matrices, Interscience Publishers, New York, 1959.

[9] N. Gastinel, "Sur les calcul des products de matrices", Numerische Mathematik, Vol. 17, 1971, pp. 222-229.

[10] F.L. Hitchcock, "The Expansion of a Tensor or a Polyadic as a Sum of Products", Journal of Mathematics and Physics, Vol. VI, pp. 164-189, 1926-27.

[11] J. Hopcroft and L. Kerr, "On Minimizing the Number of Multiplications Necessary for Matrix Multiplication", SIAM J. Appl. Math., Vol. 20, pp. 30-36, 1971.

[12] J. Hopcroft and J. Musinski, "Duality Applied to the Complexity of Matrix Multiplication and Other Bilinear Forms", Proceedings of the Fifth Annual Symposium on the Theory of Computing, Austin, Texas, 1973, pp. 73-87.

[13] Z. Kedem, "Studies in Algebraic Computational Complexity", Ph.D. Thesis, Israel Institute of Technology, Haifa, April, 1973.

[14] D. Kirkpatrick, "On the Number of Additions Required to Compute Certain Functions", Proceedings of the Fourth Annual ACM Symposium on the Theory of Computing, Denver, Colorado, 1972, pp. 94-101.

[15] D. Knuth, *The Art of Computer Programming*, Volume II: *Seminumerical Algorithms*, Addison-Wesley, 1969.

[16] S. Lang, *Algebra*, Addison-Wesley, 1967.

[17] I. Munro, "Some Results Concerning Efficient and Optimal Algorithms", Proceedings of the Third Annual ACM Symposium on the Theory of Computing, Shaker Heights, Ohio, 1971, pp. 40-44.

[18] R. Oldenburger, "On Canonical Binary Trilinear Forms", *Bulletin of the American Mathematical Society*, Volume 38, pp. 385-7, 1932.

[19] J.E. Savage, "Computational Work and Time on Finite Machines", *Journal of the Association for Computing Machinery*, Volume 19, pp. 660-674, October, 1972.

[20] A. Schönhage and V. Strassen, "Fast Multiplication of Large Numbers", *Computing*, Volume 7, pp. 281-292, 1971.

[21] V. Strassen, "Gaussian Elimination is Not Optimal", *Numerische Mathematik*, Volume 13, pp. 354-6, 1969.

[22] V. Strassen, "Evaluation of Rational Functions", in *Complexity of Computer Computations*, (R. Miller and J. Thatcher, editors), Plenum Press, 1972.

[23] V. Strassen, "Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten", *Numerische Mathematik*, Volume 20, pp. 238-251, 1973.

[24] T. Vari, "On the Number of Multiplications Required to Compute Quadratic Functions", Technical Report TR 72-120, Department of Computer Science, Cornell University, Ithaca, New York, 1972.

[25] S. Winograd, "A New Algorithm for Inner Product", *IEEE Transactions on Computers*, Volume 17, pp. 693-4, 1968.

[26] S. Winograd, "On the Number of Multiplications Necessary to Compute Certain Functions", *Communications of Pure and Applied Mathematics*, Volume 23, pp. 165-179, 1970.

[27] S. Winograd, "On the Multiplication of 2 x 2 Matrices", Linear Algebra and Applications, Volume 4, pp. 381-8, 1971.

[28] S. Winograd, "Some Remarks on Fast Multiplication of Polynomials", Proceedings of the CMU Symposium on the Complexity of Sequential and Parallel Numerical Algorithms, May, 1973.

Appendix A.1: Some Elementary and Important Characteristic Functions and Realizations

In this appendix we consider some elementary characteristic functions corresponding to various operations and study realizations of these characteristic functions. The best understood class of such functions are those which arise in polynomial multiplication problems. We denote the next appendix to a discussion of realizations of $P_{n,m}(s)$ over $\mathbb{Z}$ for various n and m. As mentioned at various places throughout the thesis, in order for any general solution to the characteristic function realization problem to be found it is necessary to understand the basic building blocks and their interconnections. We deal with only the former objective here and describe some heuristics which have been developed to handle these problems.

We begin by studying complex number multiplication for which $I(s) = \begin{bmatrix} s_1 & s_2 \\ s_2 & -s_1 \end{bmatrix}$ and seek minimal realizations of the form $I(s)=CA(s)B$. The following proof of a result due to Munro [17] and Winograd [27] is included because of its directness.

Theorem 1: $\delta_{\mathbb{R}}(I(s)) > 2$.

Proof: If $\delta_{\mathbb{R}}(I(s)) = 2$, then there exist 2x2 matrices A(s), B and C such that $I(s) = CA(s)B$, where $A(s) = \begin{bmatrix} \ell_1(s) & 0 \\ 0 & \ell_2(s) \end{bmatrix}$ for $\ell_1(s)$ and $\ell_2(s)$ linear forms in s. This implies that the determinant of I(s) can be factored as $k\ell_1(s)\ell_2(s)$ over the real field. But $\det(I(s)) = s_1^2+s_2^2$ which is irreducible over $\mathbb{R}$, a contradiction. Therefore, $\delta_{\mathbb{R}}(I(s)) > 2$.

We can now generate all degree 3 realizations of I(s) over $F_3 = \{-1,0,1\}$.

Theorem 2: If $I(s) = CA(s)B$ is a degree 3 realization over F_3, then either the realization is one of 14 standard realizations or there exist invertible diagonal matrices M_1 and M_2 over F_3 and permutations P and R such that $(CM_1P^{-1})(PM_1^{-1}A(s)M_2R^{-1})(RM_2^{-1}B)$ is a standard realization.

Proof: Let $I(s) = (a_{11}s_1+a_{12}s_2)M_1+(a_{21}s_1+a_{22}s_2)M_2+(a_{31}s_1+a_{32}s_2)M_3$ where M_1, M_2 and M_3 are rank one matrices. It is clear that any two rows of the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

must be linearly independent or a specification of s_1 and s_2 would result in a contradiction. Therefore modulo sign changes and permutations, there are four choices of A(s) as given below.

Remark: The 14 realizations of I(s) over F_3 are given by

$$\mathscr{R}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathscr{R}_8 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\mathscr{R}_2 = \begin{bmatrix} 0 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \mathscr{R}_9 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\mathscr{R}_3 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} \qquad \mathscr{R}_{10} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathscr{R}_4 = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \qquad \mathscr{R}_{11} = \begin{bmatrix} 0 & 1 & 1 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathscr{R}_5 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \qquad \mathscr{R}_{12} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathscr{R}_6 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \qquad \mathscr{R}_{13} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}$$

$$\mathscr{R}_7 = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \qquad \mathscr{R}_{14} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_2 & & \\ & s_1 & \\ & & s_1+s_2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$$

and are related as shown below where T denotes similarity under transpose, PE denotes similarity under permutation equivalence and S denotes similarity under $\Sigma_{F_3}(I(s))$ the stabilizer set.

$$\begin{tikzcd}
10 \arrow[r, leftrightarrow, "T"] & 3 \arrow[d, leftrightarrow, "PE"] \\
6 \arrow[r, leftrightarrow, "T"] \arrow[d, leftrightarrow, "S"] & 13 \\
14 \arrow[r, leftrightarrow, "T"] & 7 \arrow[d, leftrightarrow, "PE"] \\
12 \arrow[r, leftrightarrow, "T"] \arrow[d, leftrightarrow, "PE"] & 5 \\
1 \arrow[r, leftrightarrow, "T"] \arrow[d, leftrightarrow, "S"] & 8 \\
2 \arrow[r, leftrightarrow, "T"] & 9
\end{tikzcd}$$

Next we study quaternion multiplication for which the characteristic function is

$$J(s) = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ s_2 & -s_1 & s_4 & -s_3 \\ s_3 & -s_4 & -s_1 & s_2 \\ s_4 & s_3 & -s_2 & -s_1 \end{bmatrix}.$$

By expanding the results used to show $\delta(I(s)) = 3$, we can prove

Theorem 3: i) $\delta_{\mathbb{R}}(J(s)) \geqslant 7$

ii) $\delta_{\mathbb{Z}}(J(s)) \geqslant 8$

iii) $\delta_K(J(s)) \leqslant 8$ if K is any subset of $\mathbb{R}$ containing $\{1,-1,0,\frac{1}{2},\frac{1}{4},2\}$

Proof: i) follows from the argument given in the proof of Theorem 2 and an extension of this argument to study the nature of the linear forms yields ii). iii) follows from the factorization $J(s) = CA(s)B$ where

$$C = \begin{bmatrix} -1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & -1 & -1 & 0 & 0 & 1 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad B^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$A(s) = \begin{bmatrix} \frac{s_1-s_2-s_3-s_4}{4} & & & & & & & \\ & \frac{s_2-s_1-s_3-s_4}{4} & & & & & & \\ & & \frac{s_3-s_1-s_2-s_4}{4} & & & & & \\ & & & \frac{s_4-s_1-s_2-s_3}{4} & & & & \\ & & & & -2s_1 & & & \\ & & & & & -2s_2 & & \\ & & & & & & -2s_3 & \\ & & & & & & & -2s_4 \end{bmatrix}$$

An upper bound of 10 on $\delta_{F_3}(J(s))$ was obtained by Fiduccia [6] using a construction which can be extended to show that $\delta_{F_3}(K_n(s)) = \delta_{\mathbb{Z}}(K_n(s)) = \frac{n(n+1)}{2}$ where $K_n(s)$ is an nxn symmetric characteristic function of index $\frac{n(n+1)}{2}$. An nxn symmetric (or asymmetric) characteristic function containing p different entries above the diagonal and q different entries on the diagonal is of degree between p+q and p+n. However, when repeated entries occur the gap between these bounds grows.

We can study an interesting class of asymmetric characteristic functions by studying the set $\{L_n(s)\}$ of characteristic functions arising from Lie Bracket computations on nxn matrices. Observing that the Lie bracket of nxn matrices is of zero trace reduces $L_n(s)$ to an $n^2-1 \times n^2-1$ characteristic function of index n^2-1. We can express $L_n(s)$ as the last n^2-1 rows and columns of $M_n(s)-M_n^T(s)$ reduced to a nondegenerate form by eliminating one indeterminate. It is clear that $\delta(L_n(s)) \leqslant 2\delta(M_n(s))$ and the following results hold.

<u>Theorem 4</u>: i) $\delta_K(L_2(s)) = 5 \quad$ if $F_3 \subset K \subset \mathbb{R}$

ii) $\delta_{F_3}(L_4(s)) \leqslant 70$

iii) $\delta_{F_3}(L_2n(s)) \leqslant 2\cdot 7^n - 7\cdot 2^n$

<u>Proof</u>: i) $L_2(s) = \begin{bmatrix} 0 & s_1 & s_2 \\ -s_1 & 0 & s_3 \\ -s_2 & -s_3 & 0 \end{bmatrix}$ which is also the characteristic function for vector cross product calculations. To show that $\delta(L_2(s)) > 4$, we observe that $\delta(L_2(s)) > 3$ and that if $\delta(L_2(s)) = 4$, then there is a characteristic function $M(s)$ of degree 1 such that $L_2(s)-M(s)$ is of degree 3. If $L_2(s)-M(s)$ is of degree 3, then there exist 3 specifications

of the triple $(s_1\ s_2\ s_3)$ which are linearly independent and such that $L_2(s)-M(s)$ evaluated at these points is of rank one. A detailed calculation yields a contradiction and hence $\delta(L_2(s)) \geqslant 5$. A factorization of dimension 5 over F_3 is given by

$$L_2(s) = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} s_1 & & & & \\ & s_1+s_2 & & & \\ & & s_3-s_2 & & \\ & & & s_3 & \\ & & & & s_2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

We describe the algorithm for $L_4(s)$ computationally. Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = [A,B]$$

where A_{ij}, B_{ij}, C_{ij} are 2x2 matrices, then let

$$\begin{aligned} Z_1 &= [A_{11}+A_{22},\ B_{11}+B_{22}] & Z_7 &= [A_{12}-A_{22},\ B_{21}+B_{22}] \\ Z_2 &= [A_{21}+A_{22},\ B_{11}] & Y_1 &= (B_{12}+B_{22}-B_{11})A_{12} \\ Z_3 &= [A_{11},\ B_{12}-B_{22}] & Y_2 &= B_{12}(A_{11}-A_{12}-A_{22}) \\ Z_4 &= [A_{22},\ B_{21}-B_{11}] & Y_3 &= (B_{11}-B_{12}-B_{21}-B_{22})(A_{12}-A_{21}) \\ Z_5 &= [A_{11}+A_{12},\ B_{22}] & Y_4 &= B_{21}(-A_{11}+A_{21}+A_{22}) \\ Z_6 &= [A_{21}-A_{11},\ B_{11}+B_{12}] & Y_5 &= (B_{11}-B_{21}-B_{22})A_{21} \end{aligned}$$

$$\begin{aligned} C_{11} &= Z_1+Z_4-Z_5+Z_7+(Y_1+Y_3+Y_5) & C_{21} &= Z_2+Z_4-(Y_4+Y_5) \\ C_{12} &= Z_3+Z_5-(Y_1+Y_2) & C_{22} &= Z_1+Z_2+Z_3+Z_6+Y_1+Y_3+Y_5 \end{aligned}$$

and thus $\delta(L_4(s)) \leqslant 5\delta(M_2(s))+7\delta(L_2(s)) = 70$.

iii) By a similar recursion, $\delta(L_{2^n}(s)) \leqslant 10\ \ \delta(M_{2^{n-1}}(s))+2\delta(L_{2^{n-1}}(s))$

and since $\delta(M_{2^n}(s)) \leq 7^n$, this recursion yields that

$$\delta(L_{2^n}(s)) \leq 10\delta(M_{2^{n-1}}(s))+2\cdot 10\delta(M_{2^{n-2}}(s))+\ldots+2^{n-3}\cdot 10\delta(M_4(s))+2^{n-2}\delta(L_4(s))$$

$$\leq 10(7^{n-1}+2.7^{n-2}+\ldots+2^{n-3}7^2) + 2^{n-2}\cdot 70$$

$$= 10\left(\frac{7^{n-1}-7(2)^{n-2}}{1-2/7} \right)+ 2^{n-2}\cdot 70 = 2\cdot 7^n-7\cdot 2^n.$$

Other elementary characteristic functions of importance arise in polynomial multiplication and matrix multiplication. While we have studied these elsewhere, we present extensions here to complete the section. Strassen's algorithm [21] for 2x2 matrix multiplication and a minimal realization of $P_{3,3}(s)$ over $\mathbb{Z}$ have been presented within the body of the thesis. It is easily verified that the degree 3 realization

$$P_{2,2}(s) = CA(s)B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} s_1-s_2 & & \\ & s_2 & \\ & & s_3-s_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$ and Strassen's

algorithm can be extended to yield realizations of degrees 3^n and 7^n respectively for $P_{2^n,2^n}(s)$ and $M_{2^n,2^n,2^n}(s)$. Among realizations over $\mathbb{Z}$, these may be minimal, however it is clear that over any subfield K of $\mathbb{C}$, $\delta_K(P_{2^n,2^n}(s)) = 2\cdot 2^n-1 > 3^n$. The minimal degree of $M_{2^n,2^n,2^n}(s)$ is an open question of central importance to arithmetic complexity theory.

The results of this section represent an explanation of heuristics necessary to realize some elementary characteristic functions and to combine these to generate more intricate characteristic functions. These results are presented as the building blocks for a general scheme for realizing characteristic functions.

Appendix A.2: Lower and Upper Bounds on Polynomial Multiplication

Factorizations of $P_{n,m}(s)$ over any subfield of $\mathbb{C}$ are well understood and through the use of the fast Fourier transform (see Schonhage and Strassen [20]) and Modular Transformations(see Borodin and Moenck [2]) algorithms of order $O(n \log n)$ over $\mathbb{C}$ and $O(n \log^2 n)$ over any subfield of $\mathbb{C}$ exist. We devote this section to studying the behavior of $\delta_{\mathbb{Z}}(P_{n,m}(s))$ extending the results of the third section of Chapter 3. We will be particularly interested in studying the behavior of $\delta_{\mathbb{Z}}(P_{n,m}(s))$ for small values of n and m. The algorithms which are asymptotically optimal, do not dominate until n is of the order of 2^{10} and thus the results presented here are of practical interest.

It is easily verified by permutation equivalence that $\delta_{\mathbb{Z}}(P_{n,m}(s)) = \delta_{\mathbb{Z}}(P_{m,n}(s)) = \delta_{\mathbb{Z}}(T_{n,m}(s)) = \delta_{\mathbb{Z}}(T_{m,n}(s))$. Furthermore, as mentioned previously all of these equivalent quantities are bounded below by $N(m,n)$ and $N(n,m)$. It is thus easily verified that $\delta_{\mathbb{Z}}(P_{n,2}(s)) = \delta_{\mathbb{Z}}(P_{2,n}(s)) \geq N(2,n) = \left\lceil \frac{3n}{2} \right\rceil$ by the recursion given in Chapter 3. Furthermore, since $P_{n,2}(s) \underset{\mathbb{Z}}{\subset} R_n(t) = \bigoplus_{i=1}^{n/2} P_{2,2}(t_i)$ for n even and $R_n(t) = R_{n-1}(t) \oplus \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ for n odd, it is clear that $\delta_{\mathbb{Z}}(P_{n,2}(s)) \leq \delta(R_n(t)) = \left\lceil \frac{3n}{2} \right\rceil$. A minimal realization of $P_{n,2}(s)$ is thus given as the direct sum of $\frac{n}{2}$ minimal realizations of $P_{2,2}(s)$ (with additional terms if n is odd). A more interesting problem is the study of $\delta\ (P_{n,3}(s))$. The recursion of Chapter 3 yields a lower bound of $\left\lceil \frac{7n}{4} \right\rceil$ if $n \not\equiv 1 \pmod 4$ and $\left\lceil \frac{7n}{4} \right\rceil + 1$ if $n \equiv 1 \pmod 4$. From this we can establish that $\delta_{\mathbb{Z}}(P_{3,3}(s)) = 6$ and a minimal realization is given in Chapter 2. Further study yields the following

<u>Theorem 5</u>: $\delta_{\mathbb{Z}}(P_{4,3}(s)) = 8$.

<u>Proof</u>: It is clear that $\delta_{\mathbb{Z}}(P_{4,3}(s)) \geq 7$ from the above. We show that $\delta_{\mathbb{Z}}(T_{4,3}(t)) > 7$ and give a degree 8 realization for $P_{4,3}(s)$ ov r $\mathbb{Z}$. In order for a factorization of $T_{4,3}(t)$ of dimension 7 to exist over $\mathbb{Z}$, it is necessary that there exist 3 7-vectors $\{v_1,v_2,v_3\}$ such that $|v_i| \geq 4 \; \forall \; i$, $|v_i+v_j| \geq 4$, $\forall \; i\neq j$ $|v_1+v_2+v_3| \geq 4$, and $|v_i \cup v_j| \geq 6 \; \forall \; i,j$. The only such v_1,v_2,v_3 are given by

$$V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$ or any matrix of this form with row and column orders changed. Therefore, a degree 7 realization of $T_{4,3}(t)$ would be of the form CA(t)B where A(t) is the diagonal matrix with iith entry the same as the ith entry of $V^T \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$. By studying the form of such realizations, it can be seen that none exists. This is true, since if $T_{4,3}(t) = t_1T_1+t_2T_2+t_3T_3$, then there must exist rank one matrices $R_1,R_2,\ldots R_7$ such that

$$R_1+R_2+R_3+R_4 = T_1$$
$$R_3+R_4+R_5+R_6 = T_2$$
$$R_1+R_3+R_5+R_7 = T_3$$

By applying results on partitions it follows that R_3 must be nonzero only in the third and fourth columns and possible forms of R_3 and such that $T_{4,3}(t)+R_3(t_1+t_2+t_3)$ is of degree 7. A degree 8 realization follows from the degree 9 realization of $P_{4,4}$ given below.

Further study of $P_{3,n}(s)$ leads to strong support for the conjecture that $\delta_{\mathbb{Z}}(P_{3,n}(s))$ grows as fast as 2n. However, it has not yet been possible to establish these results for $n \geq 6$. It is clear that the methods presented here do yield a method of determining $\delta_{\mathbb{Z}}(P_{n,m}(s))$ for any proscribed small values of n and m through a finite search. Since there are cases where such results are desirable (e.g. Winograd [28] on triple precision arithmetic), these results are of a truly practical importance.

It is easy to extend Theorem 5 to obtain the following.

Corollary: $\delta_{\mathbb{Z}}(P_{4,4}(s)) = 9$.

This corollary follows since $\delta_{\mathbb{Z}}(P_{n,m}(s)) < \delta_{\mathbb{Z}}(P_{n',m'}(s))$ if $n' > n$ and $m' \geqslant m$ or $n' \geqslant n$ and $m' > m$. Thus, despite the fact that $P_{4,4}(s) \subseteq_{\mathbb{Z}} P_{2,2}(t_1) \otimes P_{2,2}(t_2)$, $\delta_{\mathbb{Z}}(P_{4,4}(s)) = [\delta_{\mathbb{Z}}(P_{2,2}(t))]^2$. This result is of importance as it leads to the following conjecture.

Conjecture: $\delta_{\mathbb{Z}}(P_{n,n}(s)) \geqslant n^{\log_2 3} \approx n^{1.58}$.

No bilinear operation has yet been proven to grow at exponent greater than 1 in the number of input parameters. This conjecture identifies what is perhaps the easiest case to work on. The conjecture is true for $n < 6$ by the results presented here. A summary of existing upper and lower bounds on $\delta_{\mathbb{Z}}(P_{n,m})$ for small n,m is given by the table in Appendix A.4.

Appendix A.3: A Theorem Proof

We prove the last two parts of Theorem 11 of Chapter 3 here.

Statement ii) $r(3,3,3) = 5$

Proof: Let $G(s) = G_1s_1+G_2s_2+G_3s_3$ and consider 2 cases, first that G_1 is of rank 3 and second that all G_i are of rank 2 and no linear combination of the three is of rank 3. The first case gives rise to 3 subcases depending on the Jordan canonical form Λ of $G_1^{-1}G_2$. We consider here realizations of $G(s) = I_3s_1+\Lambda s_3+As_3$ since without loss of generality this form can be assumed.

Case Ia) $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$, in this case, there exists μ such that $A-\mu I$ is of rank ≤ 2, thus $\delta(G(s)) \leq \delta(I_3s_1+\Lambda s_2)+\delta((A-\mu I)s_3) \leq 3+2 = 5$.

Case Ib) $\Lambda = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{bmatrix}$ and $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. In this case, we let

$G(s) = H(s)+(\mu s_2+(a_{33}-1)s_3)H_1+s_3H_2$ where

$H_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $H_2 = \begin{bmatrix} a_{13}a_{31} & a_{13}a_{32} & a_{13} \\ a_{23}a_{31} & a_{23}a_{32} & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix}$ are of rank 1 and $H(s)$ is of degree $\leq r(2,2,3) = 3$ and thus $\delta(G(s)) \leq 1+1+3 = 5$.

Case Ic) $\Lambda = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$ and $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. It must be further true that the Jordan Canonical Form for A is $\begin{bmatrix} \mu & 1 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{bmatrix}$ or Case Ia or Ib holds. $A+\alpha\Lambda+\beta I$ has this Jordan Canonical Form for all α, β only if $a_{31}=0$, $a_{21}=-a_{32}$ and either $a_{11} = a_{33}$ or $a_{21} = 0$. If $a_{21} = a_{31} = a_{32} = 0$, then

$G(s) = G_1(s)+G_2(s)$, where $G_1(s) = (s_1+ s_2+a_{11}s_3)\cdot\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}$ and $G_2(s)$ has as its first column, a column of zeroes. In this case, $\delta(G(s)) \leqslant \delta(G_1(s))+\delta(G_2(s)) \leqslant 1+r(3,2,3) = 5$. If $a_{31} = 0$, $a_{21} = -a_{32}$ and $a_{11} = a_{33}$, then there is a choice of β such that if $P = \begin{bmatrix}1&\beta&0\\0&1&\beta\\0&0&1\end{bmatrix}$ $PAP^{-1} = \begin{bmatrix}b_{11}&b_{12}&0\\b_{21}&b_{22}&b_{23}\\0&-b_{21}&b_{11}\end{bmatrix}$ and $PIP^{-1} = I$, $P\Lambda P^{-1} = \Lambda$. This choice is given by any of the roots of $\beta^3(a_{32})-\beta^2(a_{33}-a_{22})-\beta(a_{23}-a_{12})-a_{13} = 0$. Therefore, it is sufficient to show that $H(s) = Is_1+\Lambda s_2+H_3s_2$ is of degree 5 where $H_3 = PAP^{-1}-b_{11}I-b_{23}(\Lambda-\lambda I)$. The decomposition

$$H(s) = \begin{bmatrix}s_1+\lambda s_2 & s_2+b_{21}s_3 & 0\\ s_2+b_{21}s_3 & s_1+\lambda s_2 & 0\\ 0 & 0 & 0\end{bmatrix} + \begin{bmatrix}0&0&0\\-s_2&0&s_2\\0&0&s_1+\lambda s_2\end{bmatrix}+\begin{bmatrix}0&b_{12}-b_{23}-b_{21}&0\\0&b_{22}-b_{11}&0\\0&-b_{21}&0\end{bmatrix}s_3 .$$

proves that $\delta(G(s)) \leqslant 5$.

Case II: There is no linear combination of G_1, G_2 and G_3 of rank 3. It is clear that there exist invertible matrices P_1,P_2,Q_1,Q_2, R_1 and R_2 such that the last row and column of $R_1Q_1P_1G_1P_2Q_2R_2$ are zero and that a row and the corresponding column of $R_1Q_1P_1G_2P_2Q_2R_2$ and $R_1Q_1P_1G_3P_2Q_2R_2$ are zero. If the same row and column are zero in any two of these final matrices, it is clear that $\delta(G(s)) \leqslant 5$. Therefore, we need only show that for $G = \begin{bmatrix}a&b&0\\c&d&0\\0&0&0\end{bmatrix}$, $G_2 = \begin{bmatrix}e&0&f\\0&0&0\\g&0&h\end{bmatrix}$ and $G_3 = \begin{bmatrix}0&0&0\\0&k&\ell\\0&m&n\end{bmatrix}$, $G(s) = \sum_{i=1}^{3} G_is_i$ is of degree at most 5. If a or e $\neq 0$ (or similarly d or k $\neq 0$ or h or n$\neq$0), then we can decompose $G(s)$ as $G(s) = R_1s_1+R_2s_2+R_3(s)$ if a and e are both nonzero and as $G(s) = R_1(s_1+s_2)+R_4s_2+R_5(s)$ if e = 0 where $\delta(R_3(s)) \leqslant 3$,

$\delta(R_5(s)) \leqslant 3$ by the first part of this theorem and $R_1 = \begin{bmatrix} a & b & 0 \\ c & bc/a & 0 \\ 0 & 0 & 0 \end{bmatrix}$,

$R_2 = \begin{bmatrix} e & 0 & f \\ 0 & 0 & 0 \\ g & 0 & gf/e \end{bmatrix}$ and $R_4 = \begin{bmatrix} -a & -b & f \\ -c & -bc/a & cf/a \\ g & bg/a & -fg/a \end{bmatrix}$. If $a=d=e=h=k=n=0$, then

$$G(s) = \begin{bmatrix} 0 & 0 & 0 \\ c & -b\ell/f & 0 \\ 0 & 0 & 0 \end{bmatrix} s_1 + \begin{bmatrix} 0 & b & 0 \\ 0 & b\ell/f & 0 \\ 0 & 0 & 0 \end{bmatrix} (s_1-s_3) + \begin{bmatrix} 0 & 0 & f \\ 0 & 0 & 0 \\ 0 & 0 & fm/b \end{bmatrix} (s_2-s_3) +$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ g & 0 & -fm/b \end{bmatrix} s_2 + \begin{bmatrix} 0 & b & f \\ 0 & b\ell/f & \ell \\ 0 & m & fm/b \end{bmatrix} s_3$$

and therefore $\delta(G(s)) \leqslant 5$. If f orb = 0, it is trivial to find a realization of $G(s)$ of degree $\leqslant 5$.

<u>Statement iii)</u> $r(3,n,n) \leqslant 2n$

Proof: Let $G(s) = G_1 s_1 + G_2 s_2 + G_3 s_3$ where G_i is an nxn matrix. If no linear combination of the G_i is of rank n, then it is possible, as above, to decompose $G(s)$ as the sum of an n-2xn-2 characteristic function and a characteristic function of degree 6. Therefore, we can assume without loss of generality that G_1 is invertible and further that $G(s) = I_n s_1 + \Lambda s_2 + \Lambda s_3$ where Λ is in Jordan canonical form. If $\Lambda = (\lambda_{ij})$ and there exists i such that λ_{ik} and λ_{ki} are zero for $k \neq i$, then (assuming $i = n$), the decomposition $G(s) = H_1(s)+H_2(s)+H_3(s)$ yields a realization of degree $\leqslant 2+r(3,n-1,n-1)$ where

$$H_1(s) = (s_1+\lambda_{nn}s_2+(a_{nn}-1)s_3) \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 1 \end{bmatrix} \text{ and } H_2(s) =, s_3 \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{n-1,n} \\ 1 \end{bmatrix} [a_{n1} \dots a_{n,n-1} 1].$$

If no such i exists and all eigenvalues of Λ are linked, then there is a choice of α and β such that not all eigenvalues of $\alpha I_n+\beta\Lambda+A$ are linked unless A has the same structure as Λ. In the first case, we can reduce the problem to the above construction and in the second, a trivial realization is of degree $\leqslant 2n-1$.

Appendix A.4: Some Upper and Lower Bounds

This appendix is devoted to presenting tables of upper and lower bounds on some important characteristic functions. We begin with a general table of bounds and present bounds on special cases of polynomial and matrix multiplication.

K = G(s) =	Lower Bound on $\delta_K(G(s))$	Upper Bound on $\delta_K(G(s))$	Lower Bound on $\delta_K(G(s))$	Upper Bound on $\delta_K(G(s))$	Lower Bound on $\delta_K(G(s))$	Upper Bound on $\delta_K(G(s))$
$M_{n,n,n}(s)$	$3n^2 - \lfloor\frac{5n}{2}\rfloor$	$O(n^{2.81})$[1]	$3n^2-3n+1$	$O(n^{2.81})$[1]	$3n^2-3n+1$	$O(n^{2.81})$[1]
$M_{p,q,r}(s)$ $(p\leq q\leq r)$	$pq+pr+qr$ $-(p+q+\lfloor\frac{r}{2}\rfloor)$	$O(r^{2.81})$[1]	$pq+pr+qr-$ $(p+q+r)+1$	$O(r^{2.81})$[1]	$pq+pr+qr-$ $(p+q+r)+1$	$O(r^{2.81})$[1]
$P_{n,m}(s)$	$N(n,n) \simeq 3n$	$O(n^{1.58})$[2]	$2n-1$ [3]	$2n-1$ [3]	$2n-1$ [3]	$2n-1$ [3]
$P_{n,m}(s)$ $(n\geq m)$	$N(n,m)$	$O(n^{1.58})$[2]	$n+m-1$[3]	$n+m-1$ [3]	$n+m-1$ [3]	$n+m-1$ [3]
$L_2(s)$	5	5	5	5	5	5
$L_4(s)$		70		70		70
$L_{2^n}(s)$		$2\cdot 7^n-7\cdot 2^n$		$2\cdot 7^n-7\cdot 2^n$		$2\cdot 7^n-7\cdot 2^n$
$J(s)$	8	10 [4]	7	8		

Table I: Lower and Upper Bounds on the degrees of some important characteristic functions.

[1] These bounds are due to Strassen [21]
[2] These bounds follow from the work of Karatasuba (see Knuth [15, p. 259])
[3] These bounds are implicit in Schonhage-Strassen [20]
[4] This bound is due to Fiduccia [6]

		K =	$\mathbb{Z}$		$\mathbb{C}$	
p	q	r	Lower Bound on $\delta_K(M_{p,q,r}(s))$	Upper Bound on $\delta_K(M_{p,q,r}(s))$	Lower Bound on $\delta_K(M_{p,q,r}(s))$	Upper Bound on $\delta_K(M_{p,q,r}(s))$
2	2	2	7 [1]	7 [4]	7 [2]	7 [4]
2	2	r	$\lceil \frac{7r}{2} \rceil$ [1]	$\lceil \frac{7r}{2} \rceil$ [1]	3r+1	$\lceil \frac{7r}{2} \rceil$ [1]
2	3	3	15 [1]	15 [1]	14	15 [1]
2	q	r r>3	$qr+q+\lceil \frac{3r}{2} \rceil -2$	$\lceil \frac{3qr+\max(q,r)}{2} \rceil$ [1]	qr+r+q-1	$\lceil \frac{3qr+\max(q,r)}{2} \rceil$ [1]
3	3	3	20	24 [3]	19	24 [3]
4	4	4	38	49 [4]	37	49 [4]
6	6	6	93	165	91	165

Table II: Lower and Upper Bounds on $\delta_K(M_{p,q,r}(s))$ for K = $\mathbb{Z}$ and K = $\mathbb{Z}$ and small values of $p \leq q \leq r$.

[1] Due to Hopcroft and Kerr [11]

[2] Due to Winograd [27]

[3] Due to Fiduccia [6] and Hopcroft-Musinski [12]

[4] Due to Strassen [21]

		Lower Bound	Upper Bound on
n	m	on	on
		$\delta_{\mathbb{Z}}(P_{n,m})$	$\delta_{\mathbb{Z}}(P_{n,m})$
2	m	$\left\lceil \frac{3m}{2} \right\rceil$	$\left\lceil \frac{3m}{2} \right\rceil$
3	3	6	6
3	4	8	8
3	5	10	11
3	6	11	12
3	7	13	15
3	8	14	16
3	9	17	18
3	4k (k>1)	7k	8k
4	4	9	9
4	5	11	13
4	6	12	16
4	8	15	18
4	12	23	27
5	5	13	14
6	6	15	18
7	7	18	23
8	8	20	27

Table III: Lower and Upper Bounds on $\delta_{\mathbb{Z}}(P_{n,m}(s)) = \delta_{\mathbb{Z}}(P_{m,n}(s))$.