

# FAST STRUCTURED MATRIX COMPUTATIONS: TENSOR RANK AND COHN–UMANS METHOD

KE YE AND LEK-HENG LIM

**ABSTRACT.** We discuss a generalization of the Cohn–Umans method, a potent technique developed for studying the bilinear complexity of matrix multiplication by embedding matrices into an appropriate group algebra. We investigate how the Cohn–Umans method may be used for bilinear operations other than matrix multiplication, with algebras other than group algebras, and we relate it to Strassen’s tensor rank approach, the traditional framework for investigating bilinear complexity. To demonstrate the utility of the generalized method, we apply it to find the fastest algorithms for forming structured matrix-vector product, the basic operation underlying iterative algorithms for structured matrices. The structures we study include Toeplitz, Hankel, circulant, symmetric, skew-symmetric,  $f$ -circulant, block-Toeplitz-Toeplitz-block, triangular Toeplitz matrices, Toeplitz-plus-Hankel, sparse/banded/triangular. Except for the case of skew-symmetric matrices, for which we have only upper bounds, the algorithms derived using the generalized Cohn–Umans method in all other instances are the fastest possible in the sense of having minimum bilinear complexity. We also apply this framework to a few other bilinear operations including matrix-matrix, commutator, and simultaneous matrix products.

## 1. INTRODUCTION

In this article, we systematically study the design of fast, possibly fastest, algorithms for a variety of operations involving structured matrices, as measured by the *bilinear complexity* of the problem. This notion of complexity is best known for its use in quantifying the speed of matrix-matrix product and matrix inversion in the work of Strassen [26], Coppersmith–Winograd [10], Williams [32], and many others. The current record, due to Le Gall [20], for the asymptotic bilinear complexity of matrix-matrix product for unstructured matrices is  $O(n^{2.3728639})$ .

The algorithms that we study in article will be for the following operations: (1) matrix-vector product, (2) matrix-matrix product, and (3) commutator product:

$$(A, x) \mapsto Ax, \quad (A, B) \mapsto AB, \quad (A, B) \mapsto AB - BA,$$

where  $A$  and  $B$  are structured matrices and  $x$  is a vector, of appropriate dimensions so that the products are defined.

The structured matrices studied in this article include: (i) sparse (including banded and triangular), (ii) symmetric, (iii) skew-symmetric, (iv) Toeplitz, (v) Hankel, (vi) circulant, (vii)  $f$ -circulant and skew-circulant, (viii) block-Toeplitz-Toeplitz-block (BTTB) and more generally any block structured matrices with structured blocks, (ix) triangular Toeplitz and its analogues for Hankel and circulant matrices, (x) sum of Toeplitz and Hankel. We provide algorithms of optimal bilinear complexity for all except the skew-symmetric case (for which we only have upper bounds). The optimal bilinear complexity for the Toeplitz and triangular Toeplitz matrix-vector product are well-known, due to Bini and Capovani [1], but we will obtain them using a different method (generalized Cohn–Umans) that applies more generally to all classes of structured matrices discussed here.

We will examine two different approaches: the Strassen tensor rank approach [28, 29], and the Cohn–Umans group theoretic approach [8, 7], as well as the relations between them. Our study gives a generalization of the Cohn–Umans approach in two regards: a generalization from matrix-matrix product to arbitrary bilinear operations, and a generalization from (a) group algebras to arbitrary algebras including (b) cohomology rings of manifolds, (c) coordinate rings of schemes, (d)

polynomial identity rings. We will provide the equivalent of their ‘triple product property’ in these more general contexts.

We hope to convince our readers, by way of a series of constructions involving various structured matrices and various bilinear operations, that this generalization of Cohn–Umans method could allow one to systematically uncover fast algorithms, and these could in turn be shown to be the fastest possible (in terms of bilinear complexity) via arguments based on the Strassen tensor rank approach. For instance, we will see in Section 14 that the fastest possible algorithm for multiplying a symmetric matrix to a vector involves first writing the symmetric matrix as a sum of Hankel matrices of decreasing dimensions bordered by zeros. For example, a  $4 \times 4$  symmetric matrix would have to be decomposed into

$$\begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & j \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ b & c & d & g \\ c & d & g & i \\ d & g & i & j \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & e-c & f-d & 0 \\ 0 & f-d & e-c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h-g-e+c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This is highly nonobvious to us. We would not have been able to find this algorithm without employing the generalized Cohn–Umans approach.

The main focus of our article will be the *matrix-vector product* for various structured matrices since these form the fundamental building blocks of most modern iterative algorithms for problems involving structured matrices: linear systems [6, 23], least-squares problems [2, 23], eigenvalue problems [31], evaluating analytic functions with matrix arguments [13], etc. On the other hand, problems requiring matrix-matrix and product of structured matrices are relatively uncommon; one reason being that the most common structured matrices (symmetric, Toeplitz, Hankel, etc; in fact all but circulant) are not closed under matrix-matrix products.

**Why minimize multiplications?** In modern computer processors, there is no noticeable difference in the latency of addition and multiplication [14, Tables 14-1 and 15-6]. So the reader might wonder why bilinear complexity continue to be of relevance. We provide three reasons below.

The first reason is that such algorithms apply when we have matrices in place of scalars. We illustrate this with a simple example, Gauss’s method for multiplying two complex numbers [17]. Let  $a, b, c, d \in \mathbb{R}$ . Then the usual method

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc)$$

requires four real multiplications and two real additions but Gauss’s method

$$(a + ib)(c + id) = (ac - bd) + i[(a + b)(c + d) - ac - bd] \tag{1}$$

requires three real multiplications and five real additions. If the costs of addition and multiplication are roughly the same, then Gauss’s method is a poor way for multiplying complex numbers. However, the usefulness of Gauss’s method comes into view when we multiply complex *matrices*, i.e., when we do

$$(A + iB)(C + iD) = (AC - BD) + i[(A + B)(C + D) - AC - BD]$$

where  $A, B, C, D \in \mathbb{R}^{n \times n}$ . Now Gauss’s method requires three *matrix multiplications* instead of four. Addition and multiplication of scalars may well have similar computational costs but multiplication of  $n \times n$  matrices is by any measure vastly more expensive<sup>1</sup> than addition of  $n \times n$  matrices. This observation applies more generally. For example, Strassen’s algorithm for the product of  $2 \times 2$  matrices [26] only becomes practically useful when it is applied (recursively) to the product of  $2 \times 2$  *block matrices*.

A second reason is that our comparison of addition and multiplication in the first paragraph assumes the traditional measure of computational cost, i.e., time complexity, but other measures,

<sup>1</sup>Even if the exponent of matrix multiplication turns out to be 2; note that this is asymptotic.

e.g. energy consumption, number of gates, code space, etc, have become increasingly important. For instance, a multiplier requires many more gates than an adder (e.g. 2200 gates for an 18-bit multiplier versus 125 gates for an 18-bit adder [18]), which translates into more wires and transistors on a microchip and also consumes more energy.

A third reason is that while the latency of addition and multiplication are comparable on a general purpose CPU, it is important to remember that arithmetic is performed on other microchips as well, e.g. ASIC, DSP, FPGA, GPU, motion coprocessor, etc, where this may not be true. Moreover, our second reason also applies in this context.

## 2. THE STRUCTURE TENSOR OF A BILINEAR OPERATION

A *bilinear operation* is simply a bilinear map  $\beta : U \times V \rightarrow W$  where  $U, V, W$  are vector spaces over the same field, henceforth assumed to be  $\mathbb{C}$ . For example, the operation of forming a matrix-vector product is a bilinear operation  $\beta : \mathbb{C}^{m \times n} \times \mathbb{C}^n \rightarrow \mathbb{C}^m$ ,  $(A, x) \mapsto Ax$ , since

$$\beta(aA + bB, x) = a\beta(A, x) + b\beta(B, x), \quad \beta(A, ax + by) = a\beta(A, x) + b\beta(A, y).$$

Likewise for the operations of matrix-matrix product and commutator product.

A simple but central observation in the study of bilinear complexity is that every bilinear operation is characterized by a 3-tensor and that its tensor rank quantifies the complexity, as measured solely in terms of the number of multiplications, of the bilinear operation. We start by defining this 3-tensor.

**Definition/Proposition 2.1.** Let  $\beta : U \times V \rightarrow W$  be a bilinear map. Then there exists a unique tensor  $\mu_\beta \in U^* \otimes V^* \otimes W$  such that given any  $(u, v) \in U \times V$  we have

$$\beta(u, v) = \mu_\beta(u, v, \cdot) \in W.$$

We call  $\mu_\beta$  the *structure tensor* of the bilinear map  $\beta$ .

By the definition of tensor product, there is a one-to-one correspondence between the set of bilinear maps from  $U \times V$  to  $W$  and the set of linear maps from  $U \otimes V$  to  $W$ . Therefore we do not distinguish between a bilinear map  $\beta : U \times V \rightarrow W$  and its corresponding linear map  $\beta : U \otimes V \rightarrow W$  (and denote both by  $\beta$ ).

In the special case when  $U = V = W = \mathcal{A}$  is an algebra and the bilinear map  $\beta : \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$ ,  $(u, v) \mapsto uv$ , is multiplication in  $\mathcal{A}$ . The structure tensor of  $\beta$  is called the *structure tensor of the algebra*  $\mathcal{A}$ , and is denoted by  $\mu_{\mathcal{A}}$ .

**Example 2.2** (Lie algebras). Let  $\mathfrak{g}$  be a complex Lie algebra of dimension  $n$  and let  $\{e_1, \dots, e_n\}$  be a basis of  $\mathfrak{g}$ . Let  $\{e_1^*, \dots, e_n^*\}$  be the corresponding dual basis defined in the usual way as

$$e_i^*(e_j) = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases}$$

for all  $i, j = 1, \dots, n$ . Then for each pair  $i, j \in \{1, \dots, n\}$ ,

$$[e_i, e_j] = \sum_{k=1}^n c_{ij}^k e_k,$$

for some constant numbers  $c_{ij}^k \in \mathbb{C}$ . The structure tensor of the Lie algebra  $\mathfrak{g}$  is

$$\mu_{\mathfrak{g}} = \sum_{i,j,k=1}^n c_{i,j}^k e_i^* \otimes e_j^* \otimes e_k \in \mathfrak{g}^* \otimes \mathfrak{g}^* \otimes \mathfrak{g}.$$

The constants  $c_{ij}^k$  are often called the *structure constants* of the Lie algebra and the hypermatrix [21]

$$[c_{ij}^k] \in \mathbb{C}^{n \times n \times n}$$

is the coordinate representation of  $\mu_{\mathfrak{A}}$  with respect to the basis  $\{e_1, \dots, e_n\}$ .

For a specific example, take  $\mathfrak{g} = \mathfrak{so}_3$ , the Lie algebra of real  $3 \times 3$  skew-symmetric matrices and consider the basis of  $\mathfrak{g}$  comprising

$$e_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

with dual  $e_1^*, e_2^*, e_3^*$ . Then the structure tensor of  $\mathfrak{so}_3$  is

$$\mu_{\mathfrak{so}_3} = \sum_{i,j,k=1}^3 \epsilon_{ij}^k e_i^* \otimes e_j^* \otimes e_k,$$

where

$$\epsilon_{ij}^k = \frac{(i-j)(j-k)(k-i)}{2},$$

often called the Levi-Civita symbol.

**Example 2.3** (Matrix multiplication). Consider the usual matrix product  $\beta : \mathbb{C}^{m \times n} \times \mathbb{C}^{n \times p} \rightarrow \mathbb{C}^{m \times p}$ ,  $(A, B) \mapsto AB$ . We let  $E_{ij}$  be the elementary matrix with one in the  $(i, j)$ th entry and zeros elsewhere and  $E_{ij}^*$  be the dual. Then the structure tensor for this bilinear operation is

$$\mu_{m,n,p} = \sum_{i,j,k=1}^{m,n,p} E_{ij}^* \otimes E_{jk}^* \otimes E_{ik},$$

the famous *Strassen matrix multiplication tensor*. With respect to these bases,  $\mu_{m,n,p}$  is an  $mn \times np \times mp$  hypermatrix whose entries are all zeros and ones. When  $m = n = p$ , this becomes the structure tensor of the matrix algebra  $\mathbb{C}^{n \times n}$ .

**Example 2.4** (Matrix-vector multiplication). Consider the bilinear map

$$\beta : \mathbb{C}^{n \times n} \times \mathbb{C}^n \rightarrow \mathbb{C}^n, \quad (A, x) \mapsto Ax.$$

Let  $\{E_{ij} \in \mathbb{C}^{n \times n} : i, j = 1, \dots, n\}$  and  $\{e_i \in \mathbb{C}^n : i = 1, \dots, n\}$  be the standard bases for  $\mathbb{C}^{n \times n}$  and  $\mathbb{C}^n$  respectively. Then the structure tensor of  $\beta$  is

$$\mu_{\beta} = \sum_{i,j=1}^n E_{ij}^* \otimes e_j^* \otimes e_i.$$

With respect to these bases,  $\mu_{\beta}$  is an  $n^2 \times n \times n$  hypermatrix whose entries are all zeros and ones. This is of course nothing more than a special case of the previous example with  $m = n$  and  $p = 1$ .

A comment is in order for those who are not familiar with multilinear algebra and wonders about the difference between  $\beta$  and  $\mu_{\beta}$ . Given a bilinear map  $\beta : U \times V \rightarrow W$  there exists a unique trilinear function  $\tilde{\beta} : U \times V \times W^* \rightarrow \mathbb{C}$  such that given any  $(u, v, \omega) \in U \times V \times W^*$  we have

$$\omega(\beta(u, v)) = \tilde{\beta}(u, v, \omega).$$

Furthermore, both  $\beta$  and  $\tilde{\beta}$  correspond to the same tensor  $\mu_{\beta} \in U^* \otimes V^* \otimes W$  and so  $\mu_{\beta}$  quantifies both the bilinear operation  $\beta$  and the trilinear operation  $\tilde{\beta}$ . As a concrete example, consider Example 2.4 where  $\beta : \mathbb{C}^{n \times n} \times \mathbb{C}^n \rightarrow \mathbb{C}^n$  is the matrix-vector product

$$\beta(A, x) = Ax.$$

Then  $\tilde{\beta} : \mathbb{C}^{n \times n} \times \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$  is

$$\tilde{\beta}(A, x, y) = y^{\top} Ax,$$

and they correspond to the same tensor  $\mu_{\beta} \in (\mathbb{C}^{n \times n})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$ .

We conclude this section with the simplest example, but worked out in full details for the benefit of readers unfamiliar with multilinear algebra.

**Example 2.5** (Complex number multiplication). Complex numbers form a two-dimensional algebra over  $\mathbb{R}$  and the multiplication of complex numbers is an  $\mathbb{R}$ -bilinear map

$$\beta : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}, \quad (a + bi, c + di) \mapsto (ac - bd) + (ad + bc)i,$$

for any  $a, b, c, d \in \mathbb{R}$ . Let  $e_1 = 1 + 0i = 1$  and  $e_2 = 0 + 1i = i$  be the standard basis of  $\mathbb{C}$  over  $\mathbb{R}$  and let  $e_1^*, e_2^*$  be the corresponding dual basis. The structure tensor of  $\mathbb{C}$  is, by definition, the structure tensor of  $\beta$  and is given by

$$\mu_{\mathbb{C}} = \mu_{\beta} = e_1^* \otimes e_1^* \otimes e_1 - e_2^* \otimes e_2^* \otimes e_1 + e_1^* \otimes e_2^* \otimes e_2 + e_2^* \otimes e_1^* \otimes e_2, \quad (2)$$

or, as a hypermatrix with respect to these bases,

$$\mu_{\mathbb{C}} = \begin{bmatrix} 1 & 0 & | & 0 & 1 \\ 0 & -1 & | & 1 & 0 \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}.$$

We provide here a step-by-step verification that  $\mu_{\mathbb{C}}$  is indeed the structure tensor for complex number multiplication over  $\mathbb{R}$ . Given two complex numbers  $z_1 = a + bi$ ,  $z_2 = c + di \in \mathbb{C}$ , we write them as  $z_1 = ae_1 + be_2$ ,  $z_2 = ce_1 + de_2$ . Then

$$\begin{aligned} \mu_{\mathbb{C}}(z_1, z_2) &= [(e_1^*(z_1)e_1^*(z_2) - e_2^*(z_1)e_2^*(z_2))]e_1 + [e_1^*(z_1)e_2^*(z_2) + e_2^*(z_1)e_1^*(z_2)]e_2 \\ &= [(e_1^*(ae_1 + be_2)e_1^*(ce_1 + de_2) - e_2^*(ae_1 + be_2)e_2^*(ce_1 + de_2))]e_1 \\ &\quad + [e_1^*(ae_1 + be_2)e_2^*(ce_1 + de_2) + e_2^*(ae_1 + be_2)e_1^*(ce_1 + de_2)]e_2 \\ &= (ac - bd)e_1 + (ad + bc)e_2 = (ac - bd, ad + bc) \\ &= (ac - bd) + (ad + bc)i. \end{aligned}$$

For the uninitiated wondering the usefulness of all these, we will see in Section 3 that the notion of tensor rank and its associated rank decomposition allow us to discover faster, possibly fastest, algorithms for various bilinear operations. For instance, the four-term decomposition in (2) gives us the usual algorithm for multiplying complex numbers but as we will see, one may in fact obtain a three-term decomposition for  $\mu_{\mathbb{C}}$ ,

$$\mu_{\mathbb{C}} = (e_1^* + e_2^*) \otimes (e_1^* + e_2^*) \otimes e_2 + e_1^* \otimes e_1^* \otimes (e_1 - e_2) - e_2^* \otimes e_2^* \otimes (e_1 + e_2).$$

This gives us Gauss's method for multiplying two complex numbers with three real multiplications that we saw in (1). We will have more to say about this example in Section 8.

### 3. TENSOR RANK, BORDER RANK, AND BILINEAR COMPLEXITY

The *Strassen tensor rank method* that we have alluded to in the introduction studies the optimal bilinear complexity of a bilinear operation by studying the rank and border rank of its structure tensor.

**Definition 3.1.** Let  $\mu_{\beta}$  be the structure tensor of a bilinear map  $\beta : U \times V \rightarrow W$ , we say that the *tensor rank* or just *rank* of  $\mu_{\beta}$  is  $r$  if  $r$  is the smallest positive integer such that there exist  $u_1^*, \dots, u_r^* \in U^*$ ,  $v_1^*, \dots, v_r^* \in V^*$ , and  $w_1, \dots, w_r \in W$  such that

$$\mu_{\beta} = \sum_{i=1}^r u_i^* \otimes v_i^* \otimes w_i.$$

We denote this by  $\text{rank}(\mu_{\beta}) = r$ . We say that the *border rank* of  $\mu_{\beta}$  is  $r$  if  $r$  is the smallest positive integer such that there exists a sequence of tensors  $\{\mu_n\}_{n=1}^{\infty}$  of rank  $r$  such that

$$\lim_{n \rightarrow \infty} \mu_n = \mu_{\beta}.$$

We denote this by  $\text{brank} \mu_{\beta} = r$ . We define the rank and border rank of the zero tensor to be zero.

Our interest in tensor rank is that it gives us exactly *the least number of multiplications* required to evaluate  $\beta(u, v)$  for arbitrary inputs  $u$  and  $v$ . This is established later in Proposition 3.6. In which case border gives the least number of multiplication required to evaluate  $\beta(u, v)$  up to arbitrarily high accuracy for arbitrary inputs  $u$  and  $v$ . The study of complexity of bilinear operations in this manner is called *bilinear complexity*, originally due to Strassen [28, 29] and has developed into its own subfield within complexity theory [5, Chapter 14].

To illustrate this, we will start with a simple analysis to show that the usual way of computing matrix-vector product has optimal bilinear complexity, i.e., computing the product of an  $m \times n$  matrix and an  $n$  vector requires  $mn$  multiplications and one cannot do better.

**Proposition 3.2.** *Let  $\beta : U \times V \rightarrow W$  be a bilinear map and suppose  $\text{span}(\mu_\beta(U \otimes V)) = W$ . Then*

$$\text{rank}(\mu_\beta) \geq \dim W.$$

*The role of  $W$  may be replaced by  $U$  or  $V$ .*

*Proof.* If not, then

$$\mu_\beta = \sum_{i=1}^r u_i^* \otimes v_i^* \otimes w_i \in U^* \otimes V^* \otimes W$$

for some integer  $r < \dim W$  and vectors  $u_i^* \in U^*, v_i^* \in V^*, w_i \in W$ . Hence

$$\text{span}(\mu_\beta(U \otimes V)) \subset \text{span}\{w_i : i = 1, \dots, r\}.$$

But this contradicts the assumption that  $\mu_\beta(U \otimes V) = W$ .  $\square$

As an immediate application of Proposition 3.2, we will show that for a general matrix, the usual way of doing matrix-vector product is already optimal, i.e., Strassen-type fast algorithms for matrix-matrix product do not exist when one of the matrices is a vector (has only one column or row).

**Corollary 3.3.** *Let  $\beta : \mathbb{C}^{m \times n} \times \mathbb{C}^n \rightarrow \mathbb{C}^m$  be the bilinear map defined by the matrix-vector product. Then*

$$\text{rank}(\mu_\beta) = mn.$$

*Proof.* It is easy to see that  $\text{rank}(\mu_\beta) \leq mn$ . On the other hand,

$$\mu_\beta(\mathbb{C}^n \otimes (\mathbb{C}^m)^*) = (\mathbb{C}^{m \times n})^*,$$

since for the matrix  $E_{ij}$  with  $(i, j)$ th entry one and zero elsewhere, we have

$$\mu_\beta(e_j \otimes e_i^*)(E_{ij}) = e_i^*(E_{ij}e_j) = 1.$$

Here  $\{e_j : j = 1, \dots, n\}$  is the standard basis of  $\mathbb{C}^n$  and  $\{e_i^* : i = 1, \dots, m\}$  is its dual basis for  $(\mathbb{C}^m)^*$ .  $\square$

This establishes our earlier claim that  $mn$  is the minimal number of required multiplications for forming a matrix-vector product. Next we show that we cannot do better than  $mn$  even if we are only interested in computing matrix-vector product up to arbitrary accuracy.

Clearly we have

$$\text{brank}(\mu_\beta) \leq \text{rank}(\mu_\beta)$$

in general but equality is attained in the following special case.

**Proposition 3.4.** *Let  $\beta : U \times V \rightarrow W$  be a bilinear map and assume  $\mu_\beta(U \otimes V) = W$  and  $\text{rank}(\mu_\beta) = \dim W \leq \dim U \dim V$ . Then*

$$\text{brank}(\mu_\beta) = \text{rank}(\mu_\beta).$$

*Proof.* Assume that  $\text{brank}(\mu_\beta) < \text{rank}(\mu_\beta)$ . Notice that we may regard  $\beta$  as a linear map

$$\beta : U \otimes V \rightarrow W.$$

Since  $\mu_\beta(U \otimes V) = W$  and  $\text{rank}(\mu_\beta) = \dim W$ , the rank of  $\beta$  as a linear map (or a matrix) is  $\dim W$ . Let  $r' = \text{brank}(\mu_\beta)$ . Then there is a sequence  $\{\mu_n\}_{n=1}^\infty$  of tensors of rank  $r'$  such that

$$\lim_{n \rightarrow \infty} \mu_n = \mu_\beta.$$

We can similarly regard  $\mu_n$  as a linear map  $\mu_n : U \otimes V \rightarrow W$ . Since  $\text{rank}(\mu_n) = r'$  we see that the rank of  $\mu_n$  as a linear map is at most  $r'$ . By the choice of the sequence  $\{\mu_n\}_{n=1}^\infty$ , we see that

$$\lim_{n \rightarrow \infty} \mu_n = \mu_\beta$$

as linear maps (or matrices). Hence the border rank of  $\mu_\beta$  as a linear map (or a matrix) is at most  $r'$ . However, for matrices, the notion of rank is the same as border rank. This contradicts the fact that  $\text{rank} \mu_\beta = \dim W > r'$ .  $\square$

We deduce that the usual way of performing matrix-vector product is also optimal even if we are only interested in approximating the result up to arbitrary accuracy.

**Corollary 3.5.** *The border rank of the structure tensor of  $m \times n$  matrix-vector product is  $mn$ .*

We now give the deferred proof establishing the role of tensor rank in bilinear complexity. This simple result is well-known and classical (cf. the discussions in [5, 27, 28, 29]) but to the best of our knowledge it has never been written down explicitly. We provide it here as a small public service.

**Proposition 3.6.** *The rank of  $\mu_\beta$  equals the least number of multiplications needed to compute the bilinear map  $\beta$ .*

*Proof.* The proof is straightforward. Given  $u \in U$  and  $v \in V$ , then by definition

$$\mu_\beta(u, v, \cdot) = \beta(u, v) \in W.$$

Since  $\text{rank}(\mu_\beta) = r$  we may write  $\mu_\beta$  as

$$\mu_\beta = \sum_{i=1}^r u_i^* \otimes v_i^* \otimes w_i$$

for some  $u_i^* \in U^*$ ,  $v_i^* \in V^*$  and  $w_i \in W$ . Hence

$$\mu_\beta(u, v) = \sum_{i=1}^r u_i^*(u) v_i^*(v) w_i \in W.$$

Notice that  $u_i^*(u)$  and  $v_i^*(v)$  are complex numbers and thus to compute  $\mu_\beta$  we only need  $r$  multiplications.  $\square$

We highlight a common pitfall in the precise meaning of the word ‘multiplication’ used in the context of bilinear complexity. Here it refers strictly to the multiplications of indeterminates but excludes multiplications of a constant and an indeterminate or of two constants. For example, we need one multiplication to calculate  $(x, y) \mapsto x \cdot y$  but zero multiplication to calculate  $x \mapsto cx$  for any constant  $c \in \mathbb{C}$ . We will use the term ‘scalar multiplication’ to refer to the multiplication of two scalars or that of a scalar and an indeterminate over the given field. For instance,  $2 \cdot 3$  or  $2 \cdot x$  each requires one scalar multiplication to compute.

So in the context of bilinear complexity, a discrete Fourier transform may be computed with just  $O(1)$ , in fact zero, multiplications. The usual  $O(n \log n)$  complexity in FFT counts scalar multiplications. As the case of FFT illustrates, one reason for the exclusion of multiplication involving constants is that this part may often be performed with specialized subroutines or implemented in

hardware. On the other hand, bilinear complexity counts only multiplications of variable quantities that could change from input to input.

In the most widely studied case of matrix multiplication  $\beta : \mathbb{C}^{n \times n} \times \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ , the rank of  $\mu_\beta$  informs us about the total arithmetic complexity of  $\beta$ , i.e., counting both additions and multiplications of indeterminates, as the following theorem from [5] shows.

**Theorem 3.7** (Strassen). *Let  $R_n$  be the rank of  $\mu_\beta$  and let  $M_n$  the computational complexity of the matrix multiplication. Then*

$$\inf\{\tau \in \mathbb{R} : M_n = O(n^\tau)\} = \inf\{\tau \in \mathbb{R} : R_n = O(n^\tau)\}.$$

This says that asymptotically, the order of rank of  $\mu_\beta$  equals the order of the computational complexity of matrix multiplication. Moreover, combined with Proposition 3.6, the number of multiplications needed in matrix multiplication dominates the number of additions. This is a very special phenomenon. In general, the number of multiplications cannot dominate the number of additions.

#### 4. TENSOR RANKS OF STRUCTURE TENSORS OF ALGEBRAS

Let  $\mathcal{A}$  be an algebra of dimension  $n$ . Let  $a_1, \dots, a_n$  be a basis of  $\mathcal{A}$  and  $a_1^*, \dots, a_n^*$  be its dual basis for  $\mathcal{A}^*$ . Recall that the structure constants  $c_{ij}^k$  determine the multiplication operation in  $\mathcal{A}$ , which we denote by  $m_{\mathcal{A}} : \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$ ,

$$m_{\mathcal{A}}(a_i, a_j) = \sum_{k=1}^n c_{ij}^k a_k, \quad i, j = 1, \dots, n.$$

The structure tensor  $\mu_{\mathcal{A}} \in \mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{A}$  is then

$$\mu_{\mathcal{A}} = \sum_{i,j,k=1}^n c_{ij}^k a_i^* \otimes a_j^* \otimes a_k.$$

Note that  $\mu_{\mathcal{A}}$  does not depend on the choice of basis and neither does the tensor and border ranks of  $\mu_{\mathcal{A}}$ .

When  $\mathcal{A} = \mathbb{C}^{n \times n}$ ,  $\mu_{\mathcal{A}} = \mu_{n,n,n}$  is the Strassen matrix multiplication tensor for product of square matrices. Inspired by the Cohn–Umans approach [8] that we will discuss in the next section, we would like to study the rank of  $\mu_{\mathcal{A}}$  for an arbitrary algebra, with a view towards embedding an operation whose bilinear complexity is difficult to analyze into an algebra where the task is easier. The first question that one needs to answer is the relation between ranks of the respective multiplication tensors.

**Proposition 4.1.** *If an algebra  $\mathcal{A}$  can be embedded into another algebra  $\mathcal{B}$ , i.e.,  $\mathcal{A}$  is isomorphic to a subalgebra of  $\mathcal{B}$ , then  $\text{rank}(\mu_{\mathcal{A}}) \leq \text{rank}(\mu_{\mathcal{B}})$ .*

*Proof.* Let  $j : \mathcal{A} \hookrightarrow \mathcal{B}$  be an embedding of  $\mathcal{A}$  into  $\mathcal{B}$  as algebras<sup>2</sup>. Then it induces a surjection  $j^* : \mathcal{B}^* \twoheadrightarrow \mathcal{A}^*$  and thus a surjection

$$j^* \otimes j^* \otimes \text{Id}_{\mathcal{B}} : \mathcal{B}^* \otimes \mathcal{B}^* \otimes \mathcal{B} \twoheadrightarrow \mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{B}.$$

Let  $\delta := j^* \otimes j^* \otimes \text{Id}_{\mathcal{B}}$ . We claim that  $\delta(\mu_{\mathcal{B}}) = \mu_{\mathcal{A}}$  and to show this, it suffices to show that

$$\delta(\mu_{\mathcal{B}})(a, a', \cdot) = \mu_{\mathcal{A}}(a, a', \cdot) = m_{\mathcal{A}}(a, a'),$$

which is obvious from the definition of  $\delta$ . Let  $r$  be the rank of  $\mu_{\mathcal{B}}$  and suppose  $\mu_{\mathcal{B}}$  has a tensor decomposition

$$\mu_{\mathcal{B}} = \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes b_{i,3}$$

<sup>2</sup>Later on in the article we will consider embedding of vector spaces into algebras.



where  $b_{i,1}^*, b_{i,2}^* \in \mathcal{B}^*$  and  $b_{i,3} \in \mathcal{B}$  for  $i = 1, \dots, r$ . For notational simplicity we identify the image of  $j$  in  $\mathcal{B}$  with  $\mathcal{A}$ , regarding  $\mathcal{A}$  as a subalgebra of  $\mathcal{B}$ . Let  $\mathcal{A}^\perp$  be a subspace of  $\mathcal{B}$  so that

$$\mathcal{B} = \mathcal{A} \oplus \mathcal{A}^\perp.$$

Then we have  $b_{i,3} = a_{i,3} + a_{i,3}^\perp$  for  $i = 1, \dots, r$  and thus

$$\mu_{\mathcal{B}} = \left( \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes a_{i,3} \right) + \left( \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes a_{i,3}^\perp \right).$$

Since  $\delta(\mu_{\mathcal{B}}) = \mu_{\mathcal{A}}$ , we conclude that

$$\delta \left( \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes a_{i,3}^\perp \right) = 0 \quad \text{and} \quad \delta \left( \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes a_{i,3} \right) = \mu_{\mathcal{A}}.$$

So we obtain an expression of  $\mu_{\mathcal{A}}$  as a sum of  $r$  rank-one terms,

$$\delta \left( \sum_{i=1}^r b_{i,1}^* \otimes b_{i,2}^* \otimes a_{i,3} \right) = \sum_{i=1}^r j^*(b_{i,1}^*) \otimes j^*(b_{i,2}^*) \otimes a_{i,3} \in \mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{A},$$

and therefore  $\text{rank}(\mu_{\mathcal{A}}) \leq \text{rank}(\mu_{\mathcal{B}})$ .  $\square$

The map  $j^* : \mathcal{B}^* \rightarrow \mathcal{A}^*$  above may be viewed as the restriction of linear forms on  $\mathcal{B}$  to  $\mathcal{A}$ .

**Corollary 4.2.** *If the algebra  $\mathbb{C}^{n \times n}$  can be embedded into an algebra  $\mathcal{B}$ , then the computational complexity of multiplying two matrices is bounded by the rank of structure tensor  $\mu_{\mathcal{B}}$  of  $\mathcal{B}$ .*

If we fix a basis for  $\mathcal{A}$  and  $\mathcal{B}$  then we can identify  $\mathcal{A}$  with its dual  $\mathcal{A}^*$  and  $\mathcal{B}$  with its dual  $\mathcal{B}^*$ . Hence we may regard  $\mu_{\mathcal{A}} \in \mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{A}$  as an element of  $\mathcal{B}^* \otimes \mathcal{B}^* \otimes \mathcal{B}$ . We would like to compare the rank of  $\mu_{\mathcal{A}}$  as an element of  $\mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{A}$  and as an element of  $\mathcal{B}^* \otimes \mathcal{B}^* \otimes \mathcal{B}$ . We denote these by  $\text{rank}_{\mathcal{A}}(\mu_{\mathcal{A}})$  and  $\text{rank}_{\mathcal{B}}(\mu_{\mathcal{A}})$  respectively. We will rely on the following proposition found in [11].

**Proposition 4.3.** *Let  $U_1, \dots, U_n$  be vector spaces and let  $U'_1, \dots, U'_n$  be linear subspaces of  $U_1, \dots, U_n$  respectively. Let  $T \in U'_1 \otimes \dots \otimes U'_n$ . Suppose  $T$  has rank  $r'$  as an element in  $U'_1 \otimes \dots \otimes U'_n$  and has rank  $r$  as an element in  $U_1 \otimes \dots \otimes U_n$ , then  $r = r'$ .*

**Corollary 4.4.** *If an algebra  $\mathcal{A}$  can be embedded into another algebra  $\mathcal{B}$ , then*

$$\text{rank}_{\mathcal{A}}(\mu_{\mathcal{A}}) = \text{rank}_{\mathcal{B}}(\mu_{\mathcal{A}}).$$

Even though rank stays unchanged under embedding of vector spaces over the same ground field, the same tensor may well have different ranks over different fields. For example [3],

$$T = e_0 \otimes (e_0 \otimes e_0 - e_1 \otimes e_1) + e_1 \otimes (e_0 \otimes e_1 + e_1 \otimes e_0),$$

has rank three (over  $\mathbb{R}$ ) when viewed as an element of  $(\mathbb{R}^2)^{\otimes 3}$ , but it has rank two (over  $\mathbb{C}$ ) when viewed as an element of  $(\mathbb{C}^2)^{\otimes 3}$ .

We state here a result of Winograd [17, 33], rephrased slightly differently in terms of structure tensors of algebras.

**Theorem 4.5** (Winograd). *Let  $p(x)$  be a monic polynomial of degree  $n$  whose complete factorization over a given infinite field  $k$  is*

$$p(x) = p_1(x)^{e_1} \dots p_q(x)^{e_q}.$$

*Then the rank of the structure tensor of the algebra  $k[x]/(p(x))$  is  $2n - q$ .*

## 5. GENERALIZED COHN–UMANS METHOD AND TENSOR RANK

A major advance in the study of bilinear complexity of matrix multiplication is the *Cohn–Umans group theoretic method* proposed in [8]. The gist of idea is that one may compute multiplication in the matrix algebra by ‘embedding’ it in a judiciously chosen group algebra. The embedding in the Cohn–Umans method is however not an embedding of algebras as in Section 4, but an embedding of vector spaces. As such one needs a certain ‘triple product property’ to hold to ensure that the entries of the matrix product may still be read off from the entries of the element in the group algebra. In this section, we will generalize the Cohn–Umans method and relate it to tensor rank.

We start by briefly summarizing the Cohn–Umans method. We assume working over  $\mathbb{C}$  but the discussions in this and the next section hold for any field. Let  $G$  be a finite group and let  $\mathbb{C}[G]$  denote its group algebra over  $\mathbb{C}$ .

**Theorem 5.1** (Cohn–Umans). *Suppose that  $G$  contains subsets  $S, T, U$  of cardinality  $m, n, p$  respectively such that for any  $s, s' \in S$ ,  $t, t' \in T$ ,  $u, u' \in U$ ,  $stu = s't'u'$  implies  $s = s'$ ,  $t = t'$ ,  $u = u'$ . Then for matrices  $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ ,  $B = (b_{ij}) \in \mathbb{C}^{n \times p}$  we can associate elements  $\widehat{A}, \widehat{B} \in \mathbb{C}[G]$  as follows:*

$$\widehat{A} = \sum_{i,j=1}^{m,n} a_{ij} s_i t_j^{-1}, \quad \widehat{B} = \sum_{i,j=1}^{n,p} b_{ij} t_i u_j^{-1}. \quad (3)$$

The  $(i, j)$ th entry of  $AB$  is the coefficient of  $s_i u_j^{-1}$  in  $\widehat{A} \cdot \widehat{B}$ .

The condition in the first sentence of the theorem is called the *triple product property*. If such a condition is met, we say that  $G$  realizes  $\langle m, n, p \rangle$ .

The first step towards generalizing the Cohn–Umans method is to view the triple product property in an alternative manner, namely, it is equivalent (cf. Example 5.4) to saying that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{C}^{m \times n} \otimes \mathbb{C}^{n \times p} & \xrightarrow{j} & \mathbb{C}[G] \otimes \mathbb{C}[G] \\ m \downarrow & & \downarrow m_G \\ \mathbb{C}^{m \times p} & \xleftarrow{\text{proj}} & \mathbb{C}[G] \end{array} \quad (4)$$

Here  $j$  is the embedding of vector spaces defined by (3),  $\text{proj}$  is the projection map reading off entries of  $AB$  from the coefficients of  $\widehat{A} \cdot \widehat{B}$ , and  $m$  and  $m_G$  are respectively the multiplications of matrices and elements in the group algebra.

To be more precise about the projection map, the translation of matrix multiplication into the multiplication elements in  $\mathbb{C}[G]$  via (3) introduces some ‘junk terms’, i.e.,  $\widehat{A} \cdot \widehat{B}$  contains many coefficients that are not needed for obtaining the entries of  $AB$ . The projection map is simply a map that picks up the relevant coefficients. As will see in Section 17, there are occasions when that those ‘junk terms’ could turn out to be useful.

Another important feature of the Cohn–Umans method is that by Wedderburn Theorem,  $\mathbb{C}[G]$  can be identified with a direct sum of matrix algebras of smaller sizes determined by the irreducible representations of  $G$ , giving an efficient way to compute the product  $\widehat{A} \cdot \widehat{B}$ . Since Wedderburn Theorem holds not just for group algebras but for any semisimple algebras, one may in principle use any semisimple algebra  $\mathcal{A}$  in place of  $\mathbb{C}[G]$ . However, motivated by later examples, we will not insist that  $\mathcal{A}$  be semisimple — there will be occasions when it is useful to allow  $\mathcal{A}$  to have nilpotent elements.

The commutative diagram view of the triple product property (4) allows us to generalize it on an abstract level to arbitrary bilinear operations. Let  $\beta : U \times V \rightarrow W$  be a bilinear map and let  $\mathcal{A}$  be an algebra. If there is an injective linear map  $j : U \otimes V \rightarrow \mathcal{A} \otimes \mathcal{A}$  and a linear map  $\text{proj} : \mathcal{A} \rightarrow W$

such that the following diagram commutes:

$$\begin{array}{ccc}
 U \otimes V & \xrightarrow{j} & \mathcal{A} \otimes \mathcal{A} \\
 \beta \downarrow & & \downarrow m_{\mathcal{A}} \\
 W & \xleftarrow{\text{proj}} & \mathcal{A}
 \end{array} \tag{5}$$

then we can translate the computation of  $\beta$  into multiplication in the algebra  $\mathcal{A}$ . If this is the case, we will say that the algebra  $\mathcal{A}$  *realizes* the bilinear map  $\beta$ .

In case the reader is wondering why  $\beta : U \times V \rightarrow W$  became  $\beta : U \otimes V \rightarrow W$  in (5), recall that we do not distinguish between a bilinear map and its corresponding linear map, as we had explained after Proposition 2.1. We will adopt this convention in the rest of the article without further elaboration.

For readers unfamiliar with such constructions, we used tensor product rather than product in diagrams like (5) because we want to preserve linear structures. If we had used product in (5), then  $\beta : U \times V \rightarrow W$  has to be a bilinear map but this does not make sense in the category of vector spaces over  $\mathbb{C}$  as morphisms in this category are linear maps (and bilinear maps are in general not linear).

Note that the embedding of  $U \otimes V$  into  $\mathcal{A} \otimes \mathcal{A}$  and the projection from  $\mathcal{A}$  onto  $W$  incur zero computational costs in the context of bilinear complexity (no multiplication involved). Hence we obtain

$$\text{rank}(\mu_{\beta}) \leq \text{rank}(\mu_{\mathcal{A}}).$$

There are occasions (cf. Sections 14 and 15) when it is useful to allow an even more general framework where the multiplication of the algebra  $\mathcal{A}$  is replaced by another bilinear map in (5).

$$\begin{array}{ccc}
 U \otimes V & \xrightarrow{j} & U' \otimes V' \\
 \beta \downarrow & & \downarrow \beta' \\
 W & \xleftarrow{\text{proj}} & W'
 \end{array} \tag{6}$$

The commutative diagrams (5) and (6) may be viewed as a *generalized triple product property*. We caution our reader that despite superficial resemblance, these commutative diagrams are unrelated to [5, Lemma 14.10] — note that the direction of the bottom arrow is reversed. Also, an embedding of  $U \otimes V \rightarrow U' \otimes V'$  may not necessarily arise from an embedding of  $U \times V \rightarrow U' \times V'$ .

We will call this the *generalized Cohn–Umans method* and we will see several concrete realizations of this abstract framework in the next few sections. In following we will establish some existential guarantees for this framework.

**Theorem 5.2.** *Let  $k$  be a field. Then every bilinear map over  $k$  can be realized by an algebra  $\mathcal{A}$  over  $k$  where the rank of the structure tensor of  $\mathcal{A}$  is equal to the rank of the structure tensor of the bilinear map.*

*Proof.* Let  $\beta : U \times V \rightarrow W$  be a bilinear map and let  $\mu_{\beta}$  be the structure tensor of  $\beta$ . Without loss of generality we may assume that  $\beta$  is surjective since otherwise we may replace  $W$  by the image of  $\beta$ . Assume that  $\text{rank}(\mu_{\beta}) = r$  and that  $\mu_{\beta}$  has a decomposition

$$\mu_{\beta} = \sum_{i=1}^r u_i^* \otimes v_i^* \otimes w_i$$

for some  $u_i^* \in U^*, v_i^* \in V^*, w_i \in W$ . Let  $\mathcal{A} = k^r$ , equipped with entrywise addition and multiplication. Consider the map

$$j : U \times V \rightarrow k^r \times k^r$$

defined by  $j(u, v) = ((u_1^*(u), \dots, u_r^*(u), v_1^*(v), \dots, v_r^*(v)))$ . Lastly define the projection map

$$\text{proj} : k^r \rightarrow W, \quad \sum_{i=1}^r x_i e_i \mapsto \sum_{i=1}^r x_i w_i,$$

where  $\{e_i : i = 1, \dots, r\}$  is the standard basis of  $k^r$ . It is clear that we have a commutative diagram

$$\begin{array}{ccc} U \otimes V & \xrightarrow{j} & k^r \otimes k^r \\ \beta \downarrow & & \downarrow m_{k^r} \\ W & \xleftarrow{\text{proj}} & k^r \end{array}$$

and hence  $\beta$  is realized by the algebra  $\mathcal{A} = k^r$ .  $\square$

**Corollary 5.3.** *If  $k$  is an algebraically closed field then every bilinear map over  $k$  can be realized by the algebra  $k[x]/(x^r - 1)$  where  $r$  is the rank of the structure tensor of the bilinear map.*

*Proof.* Apply Theorem 5.2 and Wedderburn Theorem.  $\square$

One may argue that Theorem 5.2 and Corollary 5.3 are essentially tautological and not particularly useful since the algebra  $\mathcal{A}$  involved is commutative (bearing in mind that the Cohn–Umans method becomes uninteresting when the group  $G$  is abelian). We will consider another construction that yields a noncommutative algebra  $\mathcal{A}$  for a given bilinear map  $\beta$ .

The recipe for our construction goes as follows. Let  $\{u_1, \dots, u_m\}$  and  $\{v_1, \dots, v_n\}$  be bases of  $U$  and  $V$ . Let  $\{u_1^*, \dots, u_m^*\}$  and  $\{v_1^*, \dots, v_n^*\}$  be the corresponding dual bases of  $U^*$  and  $V^*$ . We may assume that  $\beta$  is nondegenerate since otherwise we may simply replace  $U$  or  $V$  by an appropriate subspace. In which case we may write the structure tensor  $\mu_\beta \in U^* \otimes V^* \otimes W$  as

$$\mu_\beta = \sum_{(i,j) : \beta(u_i, v_j) \neq 0} u_i^* \otimes v_j^* \otimes w_{ij},$$

i.e., the sum runs over all pairs of  $(i, j)$  such that  $\beta(u_i, v_j) \neq 0$ .

Let  $k\langle x_1, \dots, x_m, y_1, \dots, y_n \rangle$  be the free algebra over  $k$  with generators  $x_1, \dots, x_m$  and  $y_1, \dots, y_n$ . We consider an ideal  $I_0$  generated by the relations

$$L(x_i y_j) \sim 0 \quad \text{if and only if} \quad L(w_{ij}) = 0,$$

where  $L$  is a linear form over  $k$  in  $mn$  variables.

Next let  $I \supset I_0$  be an ideal of  $k\langle x_1, \dots, x_m, y_1, \dots, y_n \rangle$  such that  $I$  does not contain linear forms in  $x_1, \dots, x_m$  or in  $y_1, \dots, y_n$  and whenever there is a linear form  $L$  with  $L(x_i y_j) \in I$ ,  $L(x_i y_j) \in I_0$  where at least one  $(i, j)$  is a pair of indices such that  $w_{ij} \neq 0$  in the representation of  $\mu_\beta$ . Then it is easy to verify that the above commutative diagram holds for  $\mathcal{A} = k\langle x_1, \dots, x_m, y_1, \dots, y_n \rangle / I$ , by sending  $u_i$  to  $x_i$  and  $v_j$  to  $y_j$ . The algebra  $\mathcal{A}$  is a *polynomial identity ring* or *PI rings* [22].

**Example 5.4** (Triple Product Property). Here  $U = \mathbb{C}^{m \times n}$ ,  $V = \mathbb{C}^{n \times p}$ ,  $W = \mathbb{C}^{m \times p}$ , we may take  $\{E_{ij} : i = 1, \dots, m; j = 1, \dots, n\}$ ,  $\{E_{ik} : i = 1, \dots, m; k = 1, \dots, p\}$  to be the standard basis for  $U$  and  $V$ . The structure tensor may be expressed as

$$\beta_{m,n,p} = \sum_{i,k=1}^{m,p} \left( \sum_{j=1}^n E_{ij}^* \otimes E_{jk}^* \right) \otimes E_{ik}.$$

We see that  $\{E_{ik} : i = 1, \dots, m; k = 1, \dots, p\}$  is a basis of  $W$  and so there is no linear relation among them except the trivial relation  $E_{ik} = E_{ik}$ . Hence we obtain an ideal  $I_0$  generated by  $x_{ij} y_{jk} - x_{i'j'} y_{j'k}$  for all choices of  $i, j, j', k$ . As we claim above, any ideal  $I$  such that  $I \supset I_0$  does not contain linear forms in  $x_{ij}$  or  $y_{ij}$ , thus the commutative diagram holds.

Now let us understand the triple product property in this framework. Let  $G$  be a group with subsets  $S, T, U$  satisfying the triple product property. Then it is obvious that in  $k[G]$  there is no nontrivial linear relation for elements of the form  $s_i t_j^{-1}$  and  $t_j u_k^{-1}$ . The triple product property guarantees that  $(s_i t_j^{-1})(t_k u_l^{-1}) = (s_i t_{j'}^{-1})(t_k u_l^{-1})$  if and only if  $i = i', j = j', k = k'$ . Hence for such  $G$  the commutative diagram holds.

Our construction of  $\mathcal{A}$  from the structure tensor is purely formal. Usually all we could say is that  $\mathcal{A}$  is a polynomial identity ring, which does not tell us a lot. However, in special circumstances, when the ideal  $I$  is suitably chosen, we can obtain some algebras with well-understood properties that we can exploit, as we will see in the rest of this paper.

## 6. SPARSE, BANDED, AND TRIANGULAR MATRICES

We begin our study of structured matrices with *sparse matrices*, a particularly simple case that does not require the use of Cohn–Umans method. The results are also unsurprising.

One might wonder what happens to Corollaries 3.3 and 3.5 when the matrix involved has zero entries. The answer is what one would expect — the tensor and border rank are both given by the number of nonzero entries. For any  $\Omega \subseteq \{1, \dots, m\} \times \{1, \dots, n\}$ , the set of matrices with *sparsity pattern*  $\Omega$  is

$$\mathbb{C}_\Omega^{m \times n} := \{A \in \mathbb{C}^{m \times n} : a_{ij} = 0 \text{ for all } (i, j) \notin \Omega\},$$

which is clearly a vector space of dimension  $\#\Omega$ .

**Proposition 6.1.** *Let  $\beta_\Omega : \mathbb{C}_\Omega^{m \times n} \times \mathbb{C}^n \rightarrow \mathbb{C}^m$ ,  $(A, x) \mapsto Ax$ . Let  $\mu_\Omega \in (\mathbb{C}_\Omega^{m \times n})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^m$  be the corresponding structure tensor. Then*

$$\text{brank}(\mu_\Omega) = \text{rank}(\mu_\Omega) = \#\Omega.$$

*Proof.* Since the dimension of  $\mathbb{C}_\Omega^{m \times n}$  is  $\#\Omega$  and the usual matrix-vector product costs  $\#\Omega$  multiplications, the required result follows from Propositions 3.2, 3.4, and 3.6.  $\square$

The set of  $n \times n$  *banded matrices* with upper bandwidth  $k$  and lower bandwidth  $l$  is the special case when

$$\Omega = \{(i, j) \in \{1, \dots, n\} \times \{1, \dots, n\} : k < j - i < l\}.$$

Such matrices are called *diagonal* if  $(k, l) = 0$ , *lower bidiagonal* if  $(k, l) = (0, 1)$ , *upper bidiagonal* if  $(k, l) = (1, 0)$ , *tridiagonal* if  $(k, l) = (1, 1)$ , *pentadiagonal* if  $(k, l) = (2, 2)$ , *lower triangular* if  $(k, l) = (0, n - 1)$ , *upper triangular* if  $(k, l) = (n - 1, 0)$ .

**Corollary 6.2.** *The rank and border rank of the structure tensor of  $n \times n$  banded matrix-vector product are both  $n + [(n - 1) + (n - 2) + \dots + (n - k)] + [(n - 1) + (n - 2) + \dots + (n - l)]$ .*

**Corollary 6.3.** *The rank and border rank of the structure tensor of  $n \times n$  upper (or lower) triangular matrix-vector product are both  $n(n + 1)/2$ .*

## 7. CIRCULANT MATRIX

In this section, we consider the problem of computing the product of a circulant matrix with a vector or, equivalently (as we shall see), the product of two circulant matrices. We will obtain an algorithm with optimal bilinear complexity (i.e., minimum number of multiplications). This algorithm will turn out to agree with the well-known algorithm [12, Section 4.8.2] for computing product of two circulant matrices using FFT [9] but we will derive it (i) purely from the perspective of optimal bilinear complexity and (ii) using the Cohn–Umans group theoretic method. This serves as our first example of using the Cohn–Umans method for problems other than matrix multiplication.

We begin by considering a special case of matrix multiplication where we multiply a  $1 \times n$  matrix and an  $n \times 1$  matrix. Let  $a^\top = [a_1, \dots, a_n] \in \mathbb{C}^{1 \times n}$  be a row vector and let  $b = [b_1, \dots, b_n]^\top \in \mathbb{C}^{n \times 1}$

be a column vector. It is easy to verify that the cyclic group  $C_n = \langle g \mid g^n = 1 \rangle$  realizes  $\langle 1, n, 1 \rangle$  via subsets  $S = \{1\}$ ,  $T = C_n$ ,  $U = \{1\}$  that clearly satisfy the triple product property. By (3), we have

$$\widehat{a} = \sum_{i=1}^n a_i g^i, \quad \widehat{b} = \sum_{i=1}^n b_i g^{-i}.$$

The coefficients of  $g^k$  where  $k = 0, 1, \dots, n-1$  are

$$\sum_{i=1}^n a_{i+k} b_i, \quad \text{where } a_s = a_{s'} \text{ iff } s \equiv s' \pmod{n}. \quad (7)$$

To calculate  $a^\top b$  we just need the coefficient of  $1 \in \mathbb{C}[C_n]$  but not the coefficients of the remaining  $n-1$  terms (what we called ‘junk terms’ earlier). On the other hand, if we calculate the product of two circulant matrices, then these  $n-1$  terms become useful.

Let  $\text{Circ}_n(\mathbb{C})$  be the linear space of all circulant matrices. It is well-known that  $\text{Circ}_n(\mathbb{C})$  is closed under the matrix multiplication and so is an algebra.

**Proposition 7.1.** *Let  $\beta_c : \text{Circ}_n(\mathbb{C}) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ ,  $(\text{Circ}(x), y) \mapsto \text{Circ}(x)y$  be the circulant matrix-vector product and  $\mu_c \in \text{Circ}_n(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$  be the corresponding structure tensor. Let  $\mu_C \in \text{Circ}_n(\mathbb{C})^* \otimes \text{Circ}_n(\mathbb{C})^* \otimes \text{Circ}_n(\mathbb{C})$  be the structure tensor of the algebra  $\text{Circ}_n(\mathbb{C})$ . Then*

$$\text{rank}(\mu_c) = \text{rank}(\mu_C) = n.$$

*Proof.* A circulant matrix is completely specified by its first column or first row. In particular, since the product of two circulant matrices is still circulant, the product is determined by its first column. Let  $x = [x_1, \dots, x_n]^\top \in \mathbb{C}^n$  and let  $\text{Circ}(x)$  denote the circulant matrix

$$\text{Circ}(x) = \begin{bmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ x_n & x_1 & \dots & x_{n-2} & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_3 & x_4 & \dots & x_1 & x_2 \\ x_2 & x_3 & \dots & x_n & x_1 \end{bmatrix} \in \mathbb{C}^{n \times n}.$$

Observe that to calculate the matrix-matrix product  $\text{Circ}(x)\text{Circ}(y)$  it suffices to calculate the matrix-vector product  $\text{Circ}(x)[y_1, y_n, \dots, y_2]^\top$ . This implies that the structure tensor of the algebra  $\mu_C$  can be obtained from the structure tensor  $\mu_c$  of the bilinear map  $\beta_c$  and that

$$\text{rank}(\mu_c) = \text{rank}(\mu_C).$$

To compute  $\text{rank}(\mu_c)$ , observe that for two vectors  $x, y \in \mathbb{C}^n$ , we have

$$\text{Circ}(x)y = \left[ \sum_{i=1}^n x_i y_i, \sum_{i=1}^n x_i y_{i+1}, \dots, \sum_{i=1}^n x_i y_{i+n-1} \right]^\top.$$

Here we adopt the same convention in (7) that  $y_s = y_{s'}$  iff  $s \equiv s' \pmod{n}$ . Since the entries of  $\text{Circ}(x)y$  are exactly the coefficients of the product  $\widehat{x} \cdot \widehat{y} \in \mathbb{C}[C_n]$ , where  $\widehat{x}, \widehat{y} \in \mathbb{C}[C_n]$  are obtained as in (3). Now it remains to count the number of multiplications needed to form  $\widehat{x} \cdot \widehat{y}$  in  $\mathbb{C}[C_n]$ . Since  $C_n$  is the cyclic group of order  $n$ , it has exactly  $n$  representations, all of dimension one; these are indexed by the roots of unity  $1, \omega, \dots, \omega^{n-1}$  where  $\omega = e^{2k\pi i/n}$ . Denote these representations by  $V_0, \dots, V_{n-1}$  where  $V_i \simeq \mathbb{C}$  is given by<sup>3</sup>

$$\rho_i : C_n \rightarrow \mathbb{C}, \quad g \mapsto \omega^i, \quad i = 0, \dots, n-1.$$

<sup>3</sup>We do not distinguish between an irreducible representation of  $G$  and its irreducible  $\mathbb{C}[G]$ -submodule.

On the other hand, by Wedderburn Theorem we have

$$\mathbb{C}[C_n] \simeq \bigoplus_{i=0}^{n-1} V_i^* \otimes V_i,$$

i.e., we may express elements in  $\mathbb{C}[C_n]$  as  $n \times n$  diagonal matrices. Explicitly,  $\hat{x} = \sum_{i=0}^{n-1} x_i g^i$  corresponds to the diagonal matrix

$$\text{diag} \left( \sum_{i=0}^{n-1} x_i \omega^i, \sum_{i=0}^{n-1} x_i \omega^{2i}, \dots, \sum_{i=0}^{n-1} x_i \omega^{(n-1)i}, \sum_{i=0}^{n-1} x_i \right),$$

and  $\hat{y} = \sum_{i=0}^{n-1} y_i g^{-i}$  corresponds to the diagonal matrix

$$\text{diag} \left( \sum_{i=0}^{n-1} y_i \omega^{-i}, \sum_{i=0}^{n-1} y_i \omega^{-2i}, \dots, \sum_{i=0}^{n-1} y_i \omega^{-(n-1)i}, \sum_{i=0}^{n-1} y_i \right).$$

Therefore we need  $n$  multiplications to compute  $\hat{x} \cdot \hat{y}$  and thus

$$\text{rank}(\mu_c) \leq n.$$

On the other hand, it follows from Proposition 3.2 that  $\text{rank}(\mu_C) \geq n$  since the image  $\mu_C(\text{Circ}_n(\mathbb{C}) \otimes \text{Circ}_n(\mathbb{C}))$  is the whole of  $\text{Circ}_n(\mathbb{C})$ .  $\square$

The proof of Proposition 7.1 is constructive — it gives an algorithm with optimal bilinear complexity that computes a circulant matrix-vector product or a circulant matrix-circulant matrix product using only  $n$  multiplications. In fact, this algorithm is essentially the same as the well-known algorithm for circulant matrix-vector product using Fast Fourier Transform (FFT).

A departure from usual considerations in numerical linear algebra is that we only care about the number of multiplications used in the algorithm. We minimize the number of multiplications by paying the price of using more additions. We require  $n^2$  additions to execute our algorithm if we have our input  $(\text{Circ}(x), y)$  and output  $\text{Circ}(x)y$  expressed in the standard basis  $\{e_1, \dots, e_n\}$  on  $\mathbb{C}^n$ . However, if we use the Fourier basis  $f_1, \dots, f_n$  on  $\mathbb{C}^n$ , i.e., the Discrete Fourier Transform (DFT) of  $e_1, \dots, e_n$ , then we require no addition at all to execute our algorithm.

The DFT is a linear map and so computing  $f_1, \dots, f_n$  from  $e_1, \dots, e_n$  would involve only additions and scalar multiplications. Hence the use of different bases will not change the number of multiplications needed to multiply a circulant matrix to a vector (or two circulant matrices). This agrees with our expectation — a tensor and therefore its rank do not depend on the choice of bases.

Proposition 3.4 immediately gives us the border rank analogue of Proposition 7.1. It is also straightforward to obtain the analogue of Proposition 7.1 for inversion of circulant matrices.

**Corollary 7.2.** *The border ranks of the structure tensor of the bilinear operation  $\beta_c$  and the structure tensor of the algebra  $\text{Circ}_n(\mathbb{C})$  are both  $n$ .*

$$\text{brank}(\mu_c) = \text{brank}(\mu_C) = n.$$

**Corollary 7.3.** *Let  $X = \text{Circ}(x)$  be a nonsingular circulant matrix. Then one requires just  $n$  divisions to compute its inverse  $X^{-1}$ .*

*Proof.* Let  $X = \text{Circ}(x)$  where  $x = [x_1, \dots, x_n]^T \in \mathbb{C}^n$  and let  $Y = X^{-1}$  be given by  $Y = \text{Circ}(y)$  where  $y = [y_1, \dots, y_n]^T \in \mathbb{C}^n$ . As in the proof of Proposition 7.1, their corresponding images in

$\mathbb{C}[C_n]$  are

$$\begin{aligned}\widehat{X} &= \text{diag} \left( \sum_{i=0}^{n-1} x_i \omega^i, \sum_{i=0}^{n-1} x_i \omega^{2i}, \dots, \sum_{i=0}^{n-1} x_i \omega^{(n-1)i}, \sum_{i=0}^{n-1} x_i \right), \\ \widehat{Y} &= \text{diag} \left( \sum_{i=0}^{n-1} y_i \omega^{-i}, \sum_{i=0}^{n-1} y_i \omega^{-2i}, \dots, \sum_{i=0}^{n-1} y_i \omega^{-(n-1)i}, \sum_{i=0}^{n-1} y_i \right).\end{aligned}$$

Since  $\widehat{X} \cdot \widehat{Y} = I$ , we obtain

$$\widehat{Y} = \text{diag} \left( \left( \sum_{i=0}^{n-1} x_i \omega^i \right)^{-1}, \left( \sum_{i=0}^{n-1} x_i \omega^{2i} \right)^{-1}, \dots, \left( \sum_{i=0}^{n-1} x_i \omega^{(n-1)i} \right)^{-1}, \left( \sum_{i=0}^{n-1} x_i \right)^{-1} \right).$$

Hence inverting  $X$  requires  $n$  divisions<sup>4</sup>. □

### 8. $f$ -CIRCULANT AND SKEW-CIRCULANT MATRICES

We now extend the work of the previous section to  $f$ -circulant matrices. For any  $f \in \mathbb{C}$ , an  $f$ -circulant matrix is one of the form

$$\begin{bmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ fx_n & x_1 & \dots & x_{n-2} & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ fx_3 & fx_4 & \dots & x_1 & x_2 \\ fx_2 & fx_3 & \dots & fx_n & x_1 \end{bmatrix} \in \mathbb{C}^{n \times n}.$$

We denote the vector space of  $n \times n$   $f$ -circulant matrices by  $\text{Circ}_{n,f}(\mathbb{C})$ . Evidently, a 1-circulant matrix is just a usual circulant matrix. If  $f = -1$ , an  $f$ -circulant matrix is also called a *skew-circulant matrix*.

It is straightforward to see that  $\text{Circ}_{n,f}(\mathbb{C}) \simeq \mathbb{C}[x]/(x^n - f)$  and therefore also an algebra. We may employ the same techniques we used in the case  $f = 1$  to prove the following.

**Proposition 8.1.** *The rank and border rank of the structure tensor of the  $f$ -circulant matrix-vector product over  $\mathbb{C}$  are both  $n$ . Furthermore, one can invert a nonsingular  $f$ -circulant matrix over  $\mathbb{C}$  using just  $n$  divisions.*

The proofs of these statements are near identical to those in the previous section and we will not repeat them. What we will instead investigate is an interesting special case when  $n = 2$  and  $f = -1$  but over  $\mathbb{R}$  instead of  $\mathbb{C}$ .

**Proposition 8.2.** *The rank of the structure tensor of  $2 \times 2$  skew-circulant matrix-vector product over  $\mathbb{R}$  is three.*

*Proof.* The product of a  $2 \times 2$  real skew-circulant matrix with a vector is given by

$$X = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \quad v = \begin{bmatrix} c \\ -d \end{bmatrix} \in \mathbb{R}^2, \quad Xv = \begin{bmatrix} ac - bd \\ -ad - bc \end{bmatrix} \in \mathbb{R}^2.$$

Observe that to compute  $Xv$  we require just three real multiplications:

$$M_1 = (a + b)(c + d), \quad M_2 = ac, \quad M_3 = bd$$

to obtain

$$ac - bd = M_2 - M_3, \quad -ad - bc = -(M_1 - M_2 - M_3). \tag{8}$$

---

<sup>4</sup>The reader is reminded that scalar multiplications by a constant like  $\omega^i$  are not counted in bilinear complexity.



Therefore, the rank  $r$  of the structure tensor of  $2 \times 2$  skew-circulant matrix-vector product over  $\mathbb{R}$  is at most three. We show that  $r$  cannot be two. Suppose  $r = 2$ , then there exist polynomials  $M_1, M_2$  in  $a, b, c, d$ , each costing only one multiplication to evaluate, such that

$$ac - bd = \alpha_1 M_1 + \alpha_2 M_2, \quad ad + bc = \beta_1 M_1 + \beta_2 M_2.$$

Since  $a, b, c, d$  are independent variables, we must also have that

$$\det \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix} \neq 0.$$

Hence  $M_1$  and  $M_2$  are both linear combination of  $ac - bd$  and  $ad + bc$ . In particular, there exist  $s, t \in \mathbb{R}$  such that

$$M_1 = s(ac - bd) + t(ad + bc).$$

But as  $M_1$  can only involve one multiplication, it must have the form

$$M_1 = (s_1 a + s_2 b + s_3 c + s_4 d)(t_1 a + t_2 b + t_3 c + t_4 d).$$

Without loss of generality we may assume that  $s_1 = 0$  and  $t_1 = 1$ . Then  $s_2 = 0$ ,  $s_3 = s$ , and  $s_4 = t$ . These imply that  $t_2 = t/s$  and  $t_3 = t_4 = 0$  and thus

$$M_1 = (sc + td) \left( a + \frac{t}{s} b \right),$$

giving us  $t^2/s = -s$ , a contradiction since both  $s$  and  $t$  are real.  $\square$

We show a somewhat unexpected link between Proposition 8.2 and Example 2.5.

**Corollary 8.3.** *The rank and border rank of  $\mu_{\mathbb{C}}$ , the structure tensor of  $\mathbb{C}$  as an  $\mathbb{R}$ -algebra, or equivalently, the structure tensor of the  $\mathbb{R}$ -bilinear map*

$$\beta : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}, \quad (a + bi, c + di) \mapsto (ac - bd, ad + bc),$$

*are both three.*

*Proof.* Let  $z_1 = a + bi$  and  $z_2 = c + di$ . If we identify  $\mathbb{C}$  with  $\mathbb{R}^2$ , then  $z_1 z_2 = Xv$  where  $X$  and  $v$  are as defined in Proposition 8.2. From which it is clear that the structure tensor  $\mu_{\mathbb{C}} \in (\mathbb{R}^2)^* \otimes (\mathbb{R}^2)^* \otimes \mathbb{R}^2$  has rank three over  $\mathbb{R}$ . The conclusion regarding border rank follows from [11, Theorem 7.1].  $\square$

One may check that the optimal algorithm for skew-circulant matrix-vector product in (8) is in fact the same as Gauss's method for multiplication of complex numbers (1).

## 9. TOEPLITZ MATRICES

Let  $\text{Toep}_n(\mathbb{C})$  be the vector space of  $n \times n$  Toeplitz matrices. The following result is well-known, proved in [1] using methods different from those we employ below. Our objective of including this is to provide another illustration of the generalized Cohn-Umans approach where a bilinear operation is embedded in an algebra, in this case, the algebra of circulant matrices  $\text{Circ}_{2n}(\mathbb{C})$  in Section 7.

**Theorem 9.1** (Bini-Capovani). *Let  $\beta_t : \text{Toep}_n(\mathbb{C}) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$  be the Toeplitz matrix-vector product. Let  $\mu_t \in \text{Toep}_n(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$  be the structure tensor of  $\beta_t$ . Then*

$$\text{rank}(\mu_t) = 2n - 1.$$

*Proof.* We begin by observing that there is an embedding of an  $n \times n$  Toeplitz matrix  $X_n = [x_{j-i}]$  as a subblock of a  $2n \times 2n$  circulant matrix  $C_{2n}$  as follows

$$C_{2n} = \begin{bmatrix} X_n & Y_n \\ Y_n & X_n \end{bmatrix}, \tag{9}$$

where

$$Y_n = \begin{bmatrix} y & x_{-n+1} & x_{-n+2} & \cdots & x_{-1} \\ x_{n-1} & y & x_{-n+1} & \cdots & x_{-2} \\ x_{n-2} & x_{n-1} & y & \cdots & x_{-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & x_3 & \cdots & y \end{bmatrix}$$

and  $y \in \mathbb{C}$  can be arbitrarily chosen. We will choose

$$y = - \sum_{i=-(n-1)}^{n-1} x_i.$$

and by this choice of  $y$ , we only need  $2n - 1$  multiplications to compute the  $2n \times 2n$  circulant matrix-vector product. To see this, recall that in our proof of Proposition 7.1, the multiplication of  $Y_n$  by a vector of dimension  $2n$  is computed by multiplying a pair of  $2n \times 2n$  diagonal matrices. In our case, the  $2n \times 2n$  diagonal matrix corresponding to  $Y_n$  is one whose  $(1, 1)$ th entry is zero. Hence an  $n \times n$  Toeplitz matrix-vector product can be computed with just  $2n - 1$  multiplications and so the rank of the corresponding structure tensor  $\mu_t$  is at most  $2n - 1$ . On the other hand, by Proposition 3.2,  $\text{rank}(\mu_t)$  is at least  $2n - 1$  since  $\mu_t(\mathbb{C}^n \otimes (\mathbb{C}^n)^*)$  is the whole of  $\text{Toep}_n(\mathbb{C})^*$ .  $\square$

As in Section 7, Proposition 3.4 and Theorem 9.1 together give the corresponding result for border rank.

**Corollary 9.2.** *The structure tensor of the Toeplitz matrix-vector product has border rank*

$$\text{brank}(\mu_t) = 2n - 1.$$

We embedded  $\text{Toep}_n(\mathbb{C})$  into  $\text{Circ}_{2n}(\mathbb{C}) \simeq \mathbb{C}[C_{2n}]$  and the Toeplitz matrix-vector product inherits a group theoretic interpretation via this embedding. Given  $X = [x_{j-i}] \in \text{Toep}_n(\mathbb{C})$  and a vector  $z = [z_1, \dots, z_n]^\top \in \mathbb{C}^n$ , we may explicitly construct the product  $Xz$  as follows. First construct two vectors

$$a = [x_0, x_1, \dots, x_{n-1}, y, x_{-n+1}, x_{-n+2}, \dots, x_{-1}]^\top \in \mathbb{C}^{2n}, \quad b = [z_1, \dots, z_n, 0, \dots, 0]^\top \in \mathbb{C}^{2n},$$

where  $y = - \sum_{i=-(n-1)}^{n-1} x_i$ . Notice that

$$a^\top b = \sum_{i=1}^n x_{i-1} z_i$$

is the first entry of the vector  $Xz$ . As we had observed in Section 7, the cyclic group  $C_{2n} = \langle g \mid g^{2n} = 1 \rangle$  realizes  $\langle 1, 2n, 1 \rangle$ . Hence we may construct two elements in  $\mathbb{C}[C_{2n}]$ ,  $\hat{a}$  and  $\hat{b}$  as in Section 7 by

$$\hat{a} = \sum_{i=1}^n x_{i-1} g^i + y g^{n+1} + \sum_{i=n+2}^{2n} x_{i-2n-1} g^i \in \mathbb{C}[C_{2n}], \quad \hat{b} = \sum_{i=1}^n z_i g^{-i} \in \mathbb{C}[C_{2n}].$$

It is easy to see that coefficients of  $g^{2n}, g^{2n-1}, \dots, g^{n+1}$  in  $\hat{a} \cdot \hat{b}$  give the required entries of  $Xz$ .

We have seen in Corollary 7.3 that inverting a circulant matrix can be done with just  $n$  divisions. Although we may embed  $\text{Toep}_n(\mathbb{C})$  into  $\text{Circ}_{2n}(\mathbb{C})$  via (9) and  $C_{2n}$  may be inverted with  $2n$  division, there does not seem to be a way to obtain  $X_n^{-1}$  from  $C_{2n}^{-1}$ .

Assuming that we are unaware of any of the above constructions. We will now provide an illustration of how the general recipe we introduced at the end of Section 5 may be applied systematically to discover the appropriate algebra to use in the generalized Cohn–Umans method for Toeplitz

matrix-vector product. Let  $\{T_k \in \mathbb{C}^{n \times n} : k = 1, \dots, 2n-1\}$  be the standard basis for the space of Toeplitz matrices  $\text{Toep}_n(\mathbb{C})$ , i.e., the entries of  $T_k = [t_{ij}]$  are

$$t_{ij} = \begin{cases} 0 & \text{if } j - i \neq k, \\ 1 & \text{if } j - i = k. \end{cases}$$

Let  $\{e_i \in \mathbb{C}^n : i = 1, \dots, n\}$  be the standard basis of  $\mathbb{C}^n$ . Then

$$T_k e_i = e_{n-k+i},$$

where  $e_i := 0$  whenever  $i \geq n+1$  or  $i \leq 0$ . As described in Section 5, we start from the free algebra  $\mathbb{C}\langle x_1, \dots, x_{2n-1}, y_1, \dots, y_n \rangle$  and let  $I_0$  be the ideal generated by the relations

$$x_k y_i = x_{k'} y_{i'} \quad \text{whenever } 0 \leq k - i = k' - i' \leq n - 1,$$

or, equivalently,

$$x_k y_1 = x_{k+i-1} y_i \quad \text{for all } 1 \leq i \leq n, 1 \leq k \leq 2n - i. \quad (10)$$

Next we construct an ideal  $I$  such that (i)  $I_0 \subset I$ , (ii)  $I$  does not contain any linear form in  $x_i$  or  $y_j$ , and (iii) whenever  $F = L(x_k y_i)$  where  $L$  is a linear form and at least one  $x_k y_i$  appears in  $F$  for some  $0 \leq k - i \leq n - 1$ , then  $F \in I_0$ . Without loss of generality, we may assume that  $y_1 = 1$  and so (10) simplifies to

$$x_k = x_{k+i-1} y_i \quad \text{for all } 1 \leq i \leq n, 1 \leq k \leq 2n - i.$$

A moment's thought would then lead us to taking  $x_k = x^k$  where  $x$  is such that  $x^{2n} = 1$ , and also  $y_i = x^{-i+1} = x^{2n-i+1}$  for  $i = 1, \dots, n$ . It is straightforward to check the restrictions we imposed on  $I$  are satisfied by these choices, which yield the algebra  $\mathbb{C}[x]/(x^{2n} - 1) \simeq \mathbb{C}[C_{2n}]$  that we seek.

We end this section with a brief word on Toeplitz matrix- Toeplitz matrix product  $\beta_T : \text{Toep}_n(\mathbb{C}) \times \text{Toep}_n(\mathbb{C}) \rightarrow \mathbb{C}^{n \times n}$ . Note that  $\text{Toep}_n(\mathbb{C})$  is not closed under matrix multiplication [34]. The corollary below follows from Theorem 9.1 and the fact that  $XY = [Xy_1, \dots, Xy_n]$  for  $X, Y \in \text{Toep}_n(\mathbb{C})$  where  $y_i$  is the  $i$ th column of  $Y$ .

**Corollary 9.3.** *The restriction of the matrix multiplication tensor  $\mu_{n,n,n}$  to the space  $\text{Toep}_n(\mathbb{C})$  of  $n \times n$  Toeplitz matrices, regarded as a tensor in  $\text{Toep}_n(\mathbb{C})^* \otimes \text{Toep}_n(\mathbb{C})^* \otimes \mathbb{C}^{n \times n}$ , has rank at most  $n(2n - 1)$ .*

From the perspective of iterative methods for Toeplitz matrices (both linear systems and least squares), understanding  $\beta_t$  is more important than understanding  $\beta_T$ .

## 10. HANKEL MATRICES

The results in this short section follows from those in Section 9. However we state them explicitly as these results on Hankel matrices are crucial for those on *symmetric* matrices in Section 14, which might come as a surprise.

We introduce a few notations that we will use in Section 14. Given a vector  $x = [x_0, x_1, \dots, x_{2n-1}]^T \in \mathbb{C}^{2n}$ , we let

$$\text{Hank}(x) := \begin{bmatrix} x_0 & x_1 & \dots & \dots & x_{n-2} & x_{n-1} \\ x_1 & \dots & \dots & \dots & x_{n-1} & x_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n-2} & x_{n-1} & \dots & \dots & \dots & x_{2n-2} \\ x_{n-1} & x_n & \dots & \dots & x_{2n-2} & x_{2n-1} \end{bmatrix} \in \mathbb{C}^{n \times n} \quad (11)$$

be the Hankel matrix defined by  $x$ . Let  $\text{Hank}_n(\mathbb{C})$  denote the vector space of  $n \times n$  Hankel matrices.

The corresponding results for Hankel matrices may be obtained from the ones for Toeplitz matrices essentially via the observation that  $X \in \mathbb{C}^{n \times n}$  is a Hankel matrix if and only if  $JX$  and  $XJ$  are both Toeplitz matrices. Here  $J$  is the permutation matrix

$$J := \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{C}^{n \times n}.$$

Since  $J$  is a nonsingular linear transformation,  $\mu_h$  and  $\mu_t$  must have the same rank and border rank and we obtain the following from Theorem 9.1, Corollary 9.2, and Corollary 9.3.

**Corollary 10.1.** *Let  $\mu_h \in \text{Hank}_n(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$  be the structure tensor of the Hankel matrix-vector product  $\beta_h : \text{Hank}_n(\mathbb{C}) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ . Then*

$$\text{rank}(\mu_h) = \text{brank}(\mu_h) = 2n - 1.$$

*Let  $\mu_H \in \text{Hank}_n(\mathbb{C})^* \otimes \text{Hank}_n(\mathbb{C})^* \otimes \mathbb{C}^n$  be the structure tensor of the Hankel matrix-Hankel matrix product  $\beta_H : \text{Hank}_n(\mathbb{C}) \times \text{Hank}_n(\mathbb{C}) \rightarrow \mathbb{C}^{n \times n}$ . Then*

$$\text{rank}(\mu_H) \leq n(2n - 1).$$

Since  $\text{Hank}_n(\mathbb{C}) = J \text{Toep}_n(\mathbb{C}) = \text{Toep}_n(\mathbb{C})J$ , one expects a group theoretic realization of the Hankel matrix-vector multiplication. The construction is similar to that of the Toeplitz case.

## 11. TRIANGULAR TOEPLITZ/HANKEL MATRICES

We include a discussion of triangular Toeplitz (or Hankel) matrix-vector product because the result may be somewhat unexpected — its optimal bilinear complexity is exactly the same as that of a general Toeplitz (or Hankel) matrix-vector product. The fact that half the entries are zeros cannot be exploited to reduce the number of multiplications in an algorithm. Contrast this with Corollaries 3.3 and 6.3. Our methods in this section are new but the results are not, they follow from the work of Bini and Capovani [1].

Let  $\text{Toep}_n^\Delta(\mathbb{C})$  be the linear space of  $n \times n$  upper triangular Toeplitz matrices and let  $\beta_\Delta : \text{Toep}_n^\Delta(\mathbb{C}) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ ,  $(A, v) \mapsto Av$  denote the upper triangular Toeplitz matrix-vector product. We claim that the algebra  $\mathcal{A} = \mathbb{C}[x]/(x^n)$  realizes  $\beta_\Delta$ . To see this, let

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ 0 & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_0 \end{bmatrix} \in \text{Toep}_n^\Delta(\mathbb{C}) \quad \text{and} \quad v = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{n-1} \end{bmatrix} \in \mathbb{C}^n,$$

we have

$$Av = \begin{bmatrix} a_0v_0 + a_1v_1 + \cdots + a_{n-1}v_{n-1} \\ a_0v_1 + \cdots + a_{n-2}v_{n-1} \\ \vdots \\ a_0v_{n-1} \end{bmatrix} \in \mathbb{C}^n.$$

Let  $A_0, A_1, \dots, A_{n-1}$  be the ‘obvious’ basis of  $\text{Toep}_n^\Delta(\mathbb{C})$ , i.e., the  $(i, j)$ th entry of  $A_k$  is one when  $j - i = k$  and zero otherwise. Let  $e_0, \dots, e_{n-1}$  be the standard basis of  $\mathbb{C}^n$ . We define an embedding of  $\text{Toep}_n^\Delta(\mathbb{C}) \otimes \mathbb{C}^n$  into  $\mathbb{C}[x]/(x^n) \otimes \mathbb{C}[x]/(x^n)$  taking the bases elements

$$A_i \mapsto x^i, \quad e_i \mapsto x^{n-1-i}, \quad i = 0, 1, \dots, n-1.$$

For  $A = \sum_{i=0}^{n-1} a_i A_i \in \text{Toep}_n^\Delta(\mathbb{C})$  and  $v = \sum_{i=0}^{n-1} v_i e_i \in \mathbb{C}^n$ , the images  $\widehat{A}, \widehat{v} \in \mathbb{C}[x]/(x^n)$  are given by

$$\widehat{A} = a_0 1 + a_1 x + \cdots + a_{n-1} x^{n-1}, \quad \widehat{v} = v_0 x^{n-1} + v_1 x^{n-2} + \cdots + v_{n-1} 1.$$

It is straightforward to verify that  $\mathbb{C}[x]/(x^n)$  realizes  $\beta_\Delta$ . Note that  $\mathbb{C}[x]/(x^n)$  is the cohomology ring of the complex projective space  $\mathbb{C}\mathbb{P}^{n-1}$ . In particular it contains nilpotent elements and is not semisimple.

By Theorem 4.5, the structure tensor of  $\mathbb{C}[x]/(x^n)$  has rank  $2n - 1$ , from which we may deduce the following.

**Theorem 11.1.** *Let  $\mu_\Delta \in \text{Toep}_n^\Delta(\mathbb{C})^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$  be the structure tensor of the upper triangular Toeplitz matrix-vector product  $\beta_\Delta$ . Then*

$$\text{rank}(\mu_\Delta) = 2n - 1.$$

Since  $\text{Toep}_n^\Delta(\mathbb{C})$  is a linear subspace of  $\text{Toep}_n(\mathbb{C})$ , the structure tensor of upper triangular Toeplitz matrix-vector product is a projection of the structure tensor of Toeplitz matrix-vector product. However the tensor ranks of the two structure tensors are both  $2n - 1$ .

## 12. TOEPLITZ-PLUS-HANKEL MATRICES

Let  $S_1$  and  $S_2$  be two linear subspaces of  $\mathbb{C}^{n \times n}$ . Then the set  $S_1 + S_2 = \{X_1 + X_2 \in \mathbb{C}^{n \times n} : X_1 \in S_1, X_2 \in S_2\}$  is clearly also a linear subspace. If the structure tensors of the matrix-vector product for  $S_1$  and  $S_2$  have ranks  $r_1$  and  $r_2$  respectively, one might guess that the structure tensor of the matrix-vector product for  $S_1 + S_2$  has rank  $r_1 + r_2$ . However this is not true as we will see below.

**Example 12.1.** Let  $\text{Toep}_n^\Delta(\mathbb{C})$  be the linear subspace of upper-triangular Toeplitz matrices as in Section 11. Let  $\text{Toep}_n^\Delta(\mathbb{C})^\top$  be the linear subspace of lower triangular Toeplitz matrices. Clearly,

$$\text{Toep}_n^\Delta(\mathbb{C}) + \text{Toep}_n^\Delta(\mathbb{C})^\top = \text{Toep}_n(\mathbb{C}).$$

However, by Theorems 9.1 and 11.1, the structure tensors of  $\text{Toep}_n^\Delta(\mathbb{C})$ ,  $\text{Toep}_n^\Delta(\mathbb{C})^\top$ , and  $\text{Toep}_n(\mathbb{C})$  all have the same rank  $2n - 1$ .

In the special case  $S_1 = \text{Toep}_n(\mathbb{C})$  and  $S_2 = \text{Hank}_n(\mathbb{C})$ , a matrix in  $S_1 + S_2$  is often called a *Toeplitz-plus-Hankel* matrix [23, 25]. We show that the value of its rank is one less than the naive guess.

**Proposition 12.2.** *The structure tensor of the Toeplitz-plus-Hankel matrix-vector product has rank  $4n - 3$ .*

*Proof.* Let  $E \in \mathbb{C}^{n \times n}$  be the matrix of all ones. For any  $T \in \text{Toep}_n(\mathbb{C})$  and  $H \in \text{Hank}_n(\mathbb{C})$  we have

$$T + H = (T + aE) + (H - aE)$$

and  $T + aE \in \text{Toep}_n(\mathbb{C})$ ,  $H - aE \in \text{Hank}_n(\mathbb{C})$  for all  $a \in \mathbb{C}$ . We show that we may choose an appropriate  $a \in \mathbb{C}$  so that the matrix-vector product for  $T + aE$  requires only  $2n - 2$  multiplications. As in the proof of Theorem 9.1, we may embed  $X = T + aE$  into a  $2n \times 2n$  circulant matrix

$$C_{2n} = \begin{bmatrix} X & Y \\ Y & X \end{bmatrix}$$

that corresponds to a diagonal matrix whose  $(1, 1)$ th entry is zero. We may choose  $a \in \mathbb{C}$  so that the  $(2, 2)$ th entry of this diagonal matrix is also zero. Hence the matrix-vector product with  $T + aE$  costs at most  $2n - 2$  multiplications. Combined with Corollary 10.1, we see that the structure tensor of the matrix-vector product for  $T + H$  has rank at most  $4n - 3$ . On the other hand, we may check that  $\text{Toep}_n(\mathbb{C}) + \text{Hank}_n(\mathbb{C})$  has dimension  $4n - 3$ . So by Proposition 3.2, the rank is exactly  $4n - 3$ .  $\square$

## 13. BLOCK-TOEPLITZ-TOEPLITZ-BLOCK MATRICES

One of the most common Toeplitz-like structure in numerical linear algebra is that of a *block-Toeplitz-Toeplitz-block* or BTTB matrix [15, 6, 23]. As the name suggests, these are  $nk \times nk$  matrices that are  $n \times n$  block Toeplitz matrices whose blocks are themselves  $k \times k$  Toeplitz matrices, i.e.,

$$A = \begin{bmatrix} X_0 & X_1 & \cdots & X_{n-2} & X_{n-1} \\ X_{-1} & X_0 & \cdots & X_{n-3} & X_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{2-n} & X_{3-n} & \cdots & X_0 & X_1 \\ X_{1-n} & X_{2-n} & \cdots & X_{-1} & X_0 \end{bmatrix} \in \mathbb{C}^{nk \times nk},$$

where  $X_i \in \text{Toep}_k(\mathbb{C})$  for all  $i = -(n-1), \dots, -1, 0, 1, \dots, n-1$ . We write  $\text{BTTB}_{n,k}(\mathbb{C})$  for the set of  $n \times n$  block Toeplitz matrices with  $k \times k$  Toeplitz blocks.

There is of course nothing particularly special about the Toeplitz structure — may also define block-Hankel-Hankel-block or BHHB matrices, block-circulant-circulant-block or BCCB matrices, etc. In fact we will establish a general result that holds not only for any block matrices with structured blocks but those with *multiple level block structures* (e.g. block Hankel matrices whose blocks are BTTB matrices or block BHHB matrices whose blocks are BCCB matrices).

For each  $j = 1, \dots, s$ , let  $U_{k_j}$  be a linear subspace of  $\mathbb{C}^{k_j \times k_j}$ . We define the following linear subspace

$$U_{k_1} \otimes \cdots \otimes U_{k_s} \subseteq \mathbb{C}^{k_1 \times k_1} \otimes \cdots \otimes \mathbb{C}^{k_s \times k_s}$$

where ‘ $\otimes$ ’ denotes the *Kronecker product* [30]. Note that

$$\mathbb{C}^{k_1 \times k_1} \otimes \cdots \otimes \mathbb{C}^{k_s \times k_s} = \mathbb{C}^{k_1 \cdots k_s \times k_1 \cdots k_s}.$$

In particular, the linear subspace of BTTB matrices is obtained by setting  $s = 2$ ,  $k_1 = n$ ,  $k_2 = k$ , and  $U_{k_1} = \text{Toep}_n(\mathbb{C})$ ,  $U_{k_2} = \text{Toep}_k(\mathbb{C})$ , i.e.,

$$\text{Toep}_n(\mathbb{C}) \otimes \text{Toep}_k(\mathbb{C}) = \text{BTTB}_{n,k}(\mathbb{C}).$$

For  $s = 3$  and  $U_{k_i} = \text{Toep}_{k_i}(\mathbb{C})$ ,  $i = 1, 2, 3$ , we obtain  $k_1 \times k_1$  block Toeplitz matrices whose blocks are  $k_2 k_3 \times k_2 k_3$  BTTB matrices,

$$\text{Toep}_{k_1}(\mathbb{C}) \otimes \text{Toep}_{k_2}(\mathbb{C}) \otimes \text{Toep}_{k_3}(\mathbb{C}) = \text{Toep}_{k_1}(\mathbb{C}) \otimes \text{BTTB}_{k_2, k_3}(\mathbb{C}).$$

**Lemma 13.1.** *Let  $U \subseteq \mathbb{C}^{n \times n}$  and  $V \subseteq \mathbb{C}^{k \times k}$  be linear subspaces. Let*

$$\beta_U : U \times \mathbb{C}^n \rightarrow \mathbb{C}^n, \quad \beta_V : V \times \mathbb{C}^k \rightarrow \mathbb{C}^k, \quad \beta_{U \otimes V} : (U \otimes V) \times \mathbb{C}^{nk} \rightarrow \mathbb{C}^{nk}$$

*be the corresponding matrix-vector products with respective structure tensors*

$$\mu_U \in U^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n, \quad \mu_V \in V^* \otimes (\mathbb{C}^k)^* \otimes \mathbb{C}^k, \quad \mu_{U \otimes V} \in (U \otimes V)^* \otimes (\mathbb{C}^{nk})^* \otimes \mathbb{C}^{nk}.$$

*Suppose*

$$\text{rank}(\mu_U) = \dim U, \quad \text{rank}(\mu_V) = \dim V \tag{12}$$

*and*

$$\mu_U(U \otimes \mathbb{C}^n) = \mathbb{C}^n, \quad \mu_V(V \otimes \mathbb{C}^k) = \mathbb{C}^k. \tag{13}$$

*Then*

$$\text{rank}(\mu_{U \otimes V}) = \text{rank}(\mu_U) \text{rank}(\mu_V).$$

*Proof.* It is clear that  $\text{rank}(\mu_{U \otimes V})$  is bounded above by  $\text{rank}(\mu_U) \text{rank}(\mu_V)$ . So it suffices to show that  $\text{rank}(\mu_{U \otimes V})$  is bounded below by  $\text{rank}(\mu_U) \text{rank}(\mu_V)$  but this follows from applying Proposition 3.2 to the matrix-vector product

$$(U \otimes V) \otimes \mathbb{C}^{nk} \rightarrow \mathbb{C}^{nk}. \quad \square$$

The desired result for BTTB matrices follows immediately from Theorem 9.1 and Lemma 13.1.

**Corollary 13.2.** *The rank of the structure tensor of the matrix-vector product  $\beta_{\text{BTTB}} : \text{BTTB}_{n,k}(\mathbb{C}) \times \mathbb{C}^{nk} \rightarrow \mathbb{C}^{nk}$  is  $(2k-1)(2n-1)$ .*

We state a more general version of Lemma 13.1 that applies to multilevel block structures.

**Theorem 13.3.** *For  $j = 1, \dots, s$ , let  $U_{k_j} \subseteq \mathbb{C}^{k_j \times k_j}$  be a linear subspace of Toeplitz, Hankel,  $f$ -circulant, Toeplitz-plus-Hankel, symmetric, or sparse matrices (each  $U_{k_j}$  may have a different structure). Let  $\mu_j$  be the structure tensor of the matrix-vector product  $U_{k_j} \times \mathbb{C}^{k_j} \rightarrow \mathbb{C}^{k_j}$ ,  $j = 1, \dots, s$ , and let  $\mu$  be that of  $(U_1 \otimes \dots \otimes U_s) \times \mathbb{C}^{k_1 \dots k_s} \rightarrow \mathbb{C}^{k_1 \dots k_s}$ . Then*

$$\text{rank}(\mu) = \prod_{j=1}^s \text{rank}(\mu_j).$$

*Proof.* By our discussions in the previous and later sections, the conditions (12) and (13) are met for these structured matrices. The result follows by applying Lemma 13.1 inductively.  $\square$

$f$ -circulant matrices include circulant and skew-circulant ones; sparse matrices include banded and triangular ones. Note that we have excluded skew-symmetric matrices and triangular Toeplitz matrices since they do not satisfy (13).

Next we will discuss a Cohn-Umans realization of the matrix-vector product for  $U_1 \otimes \dots \otimes U_s$ :

**Proposition 13.4.** *If the algebra  $\mathcal{A}_j$  realize of the bilinear map  $\beta : U_{k_j} \times \mathbb{C}^{k_j} \rightarrow \mathbb{C}^{k_j}$  for  $j = 1, \dots, s$ , then the tensor product  $\mathcal{A} = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_s$  realizes the Kronecker product  $U_1 \otimes \dots \otimes U_s$ .*

*Proof.* It suffices to prove the statement for  $\beta_{U \otimes V} : (U \otimes V) \times \mathbb{C}^{nk} \rightarrow \mathbb{C}^{nk}$  when  $\beta_U : U \times \mathbb{C}^n \rightarrow \mathbb{C}^n$  and  $\beta_V : V \times \mathbb{C}^k \rightarrow \mathbb{C}^k$  are realized by  $\mathcal{A}$  and  $\mathcal{B}$  respectively. But this follows from routine arguments: The embeddings  $U \hookrightarrow \mathcal{A}$  and  $V \hookrightarrow \mathcal{B}$  induce an embedding of  $U \otimes V \hookrightarrow \mathcal{A} \otimes \mathcal{B}$  and the projections of  $\mathcal{A}$  onto  $\mathbb{C}^n$  and  $\mathcal{B}$  onto  $\mathbb{C}^k$  induce a projection of  $\mathcal{A} \otimes \mathcal{B}$  onto  $\mathbb{C}^{kn}$ . The more general statement then follow from induction.  $\square$

For example the matrix-vector product  $\beta_{\text{BTTB}} : \text{BTTB}_{n,k}(\mathbb{C}) \times \mathbb{C}^{nk} \rightarrow \mathbb{C}^{nk}$  is realized by the algebra

$$\mathcal{A} = \mathbb{C}[x, y]/(x^{2k} - 1, y^{2n} - 1),$$

from which we may also deduce the rank of the structure tensor of  $\beta_{\text{BTTB}}$ .

#### 14. SYMMETRIC MATRICES

We saw in Corollaries 3.3 and 3.5 that the usual way of performing matrix-vector product is already optimal for general matrices. A natural question is: What if we require the matrix to be symmetric? This is a very common situation since many, if not most, linear systems and least squares problems that arise in practice involve symmetric coefficient matrices. Despite this, we are unaware of any previous study. We show here that the optimal bilinear complexity for symmetric matrix-vector product is  $n(2n-1)$ . Surprisingly the solution involves Hankel matrices.

We begin with the observation that every symmetric matrix may be expressed as a sum of symmetric Hankel matrices bordered by zeros. A  $2 \times 2$  symmetric matrix is already a Hankel matrix. The  $3 \times 3$  and  $4 \times 4$  cases are shown explicitly below.

$$\begin{aligned} \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} &= \begin{bmatrix} a & b & c \\ b & c & e \\ c & e & f \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & d-c & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ \begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & j \end{bmatrix} &= \begin{bmatrix} a & b & c & d \\ b & c & d & g \\ c & d & g & i \\ d & g & i & j \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & e-c & f-d & 0 \\ 0 & f-d & e-c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h-g-e+c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The generalization of this observation to  $n \times n$  symmetric matrices will be established in our proof below, and together with Corollary 10.1, be used to deduce the optimal bilinear complexity of symmetric matrix-vector product. Let  $\mathcal{S}^2(\mathbb{C}^n)$  be the space of all  $n \times n$  symmetric matrices. Let  $\beta_s : \mathcal{S}^2(\mathbb{C}^n) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$  be the bilinear map of symmetric matrix-vector product and  $\mu_s \in \mathcal{S}^2(\mathbb{C}^n)^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n$ .

**Theorem 14.1.** *The optimal bilinear complexity of symmetric matrix-vector product is  $(n+1)/2$ , i.e.,  $\text{rank}(\mu_s) = (n+1)/2$ .*

*Proof.* By Proposition 3.2, we see that  $\text{rank}(\mu_\beta) \geq \dim \mathcal{S}^2(\mathbb{C}^n) = n(n+1)/2$ . On the other hand, for a given

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{1,2} & a_{2,2} & \cdots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1,n-1} & a_{2,n-1} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{1,n} & a_{2,n} & \cdots & a_{n-1,n} & a_{n,n} \end{bmatrix} \in \mathcal{S}^2(\mathbb{C}^n)$$

and a column vector  $v = [v_1, \dots, v_n]^\top \in \mathbb{C}^n$ , we claim that  $(A, v) \mapsto Av$  can be computed as the sum of several Hankel matrix-vector products of decreasing sizes. Let

$$H_1 = \text{Hank}(a_{1,1}, \dots, a_{1,n}, a_{2,n}, \dots, a_{n,n}) \in \text{Hank}_n(\mathbb{C}),$$

notation as in (11). Then  $A - H_1$  is a symmetric matrix of the form

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{S}^2(\mathbb{C}^n),$$

where  $A_2 \in \mathcal{S}^{(n-2)}(\mathbb{C})$ . Also we notice that

$$(A - H_1)v = \begin{bmatrix} 0 \\ A_2 v^{(2)} \\ 0 \end{bmatrix}$$

where  $v^{(2)} = [v_2, \dots, v_{n-1}]^\top \in \mathbb{C}^{n-2}$ . Now we can repeat the above procedure and inductively we can prove our claim. Explicitly,

$$Av = H_1 v + \begin{bmatrix} 0 \\ H_2 v^{(2)} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ H_3 v^{(3)} \\ 0 \\ 0 \end{bmatrix} + \cdots,$$

where  $H_i \in \text{Hank}_{n-2i}(\mathbb{C})$  and  $v^{(i)} \in \mathbb{C}^{n-2i}$ ,  $i = 1, \dots, \lfloor n/2 \rfloor$ .  $H_i$  and  $v^{(i)}$  are linear in the entries of  $A$  and  $v$  respectively. By Corollary 10.1, one can compute  $H_i v^{(i)}$  in  $2(n-2i) - 1$  multiplications. Hence we obtain

$$\text{rank}(\mu_s) \leq \sum_{i=0}^{\lfloor n/2 \rfloor} [2(n-2i) - 1] = \frac{n(n+1)}{2}$$

and therefore

$$\text{rank}(\mu_s) = \frac{n(n+1)}{2}. \quad \square$$



One may also interpret the proof of Theorem 14.1 as an instance of the generalized Cohn–Umans method. We have an embedding of vector spaces

$$\mathcal{S}^2(\mathbb{C}^n)^* \otimes (\mathbb{C}^n)^* \otimes \mathbb{C}^n \hookrightarrow \bigoplus_{i=0}^{\lfloor n/2 \rfloor} \text{Hank}_{n-2i}(\mathbb{C})^* \otimes (\mathbb{C}^{n-2i})^* \otimes \mathbb{C}^{n-2i}, \quad (14)$$

and for each  $i = 0, \dots, \lfloor n/2 \rfloor$ , the bilinear map  $\text{Hank}_{n-2i}(\mathbb{C}) \times \mathbb{C}^{n-2i} \rightarrow \mathbb{C}^{n-2i}$ ,  $(H_i, v^{(i)}) \mapsto H_i v^{(i)}$  is in turn realized by the algebra  $\mathbb{C}[x]/(x^{2(n-2i)} - 1)$ . Note that the object on the right-hand side of (14) is not an algebra but only a vector space — this is an application of the commutative diagram (6).

## 15. SKEW-SYMMETRIC MATRICES

A departure from other sections in this article is that in this section we do not have the optimal bilinear complexity, only upper bounds. We first discuss the case of  $3 \times 3$  skew-symmetric matrix-vector product. Let

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix} \in \Lambda^2(\mathbb{C}^3).$$

Then the usual matrix-vector multiplication gives

$$A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ay + bz \\ -ax + cz \\ -bx - cy \end{bmatrix},$$

which costs six multiplications. So the rank of the structure tensor of the skew-symmetric matrix-vector product is at most six. We will rely on the following theorem [24, 27] for the lower bound of the border rank (hence the rank) of a special 3-tensor.

**Theorem 15.1.** *Let  $T \in U \otimes V \otimes W$  where  $\dim U = \dim V = \dim W = 3$ . Let  $u_1, u_2, u_3$  be a basis of  $U$ . If we can write  $T$  as*

$$T = u_1 \otimes X_1 + u_2 \otimes X_2 + u_3 \otimes X_3,$$

with  $X_1, X_2, X_3 \in V \otimes W$  regarded<sup>5</sup> as  $3 \times 3$  matrices and if the following block matrix is nonsingular,

$$M_T = \begin{bmatrix} 0 & X_3 & -X_2 \\ -X_3 & 0 & X_1 \\ X_2 & X_1 & 0 \end{bmatrix} \in \mathbb{C}^{9 \times 9},$$

then  $\text{brank}(T) \geq 5$ . The same result holds with  $V$  or  $W$  in the role of  $U$ .

Let  $\Lambda^2(\mathbb{C}^n)$  be the space of all  $n \times n$  skew-symmetric matrices. Note that  $\dim \Lambda^2(\mathbb{C}^3) = 3$  and we may apply Theorem 15.1 to  $T = \mu_\Lambda$ , the structure tensor of the bilinear map

$$\beta_\Lambda : \Lambda^2(\mathbb{C}^3) \times \mathbb{C}^3 \rightarrow \mathbb{C}^3, \quad \left( \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}, \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) \mapsto \begin{bmatrix} ay + bz \\ -ax + cz \\ -bx - cy \end{bmatrix}.$$

Let  $e_1, e_2, e_3$  be the standard basis of  $\mathbb{C}^3$  and

$$F_1 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad F_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

be a basis of  $\Lambda^2(\mathbb{C}^3)$ . Then we may decompose  $\mu_\Lambda$  as

$$\mu_\Lambda = F_1 \otimes (e_2 \otimes e_1 - e_1 \otimes e_2) + F_2 \otimes (e_3 \otimes e_1 - e_1 \otimes e_3) + F_3 \otimes (e_3 \otimes e_2 - e_2 \otimes e_3)$$

<sup>5</sup>The result is however coordinate independent, i.e., it does not depend on our choice of the bases.

and it is easy to verify that  $M_{\mu_\Lambda}$  is nonsingular, giving us the following.

**Proposition 15.2.** *The rank and border rank of the structure tensor of skew-matrix-vector product for  $3 \times 3$  matrices are given by*

$$\text{rank}(\mu_\Lambda) = 5 \quad \text{or} \quad 6$$

and

$$\text{brank}(\mu_\Lambda) = 5$$

respectively

Next we construct an algebra that realizes the  $3 \times 3$  skew-symmetric matrix-vector product. Our candidate is

$$\mathcal{A} = \mathbb{C} \langle x_1, x_2 \rangle / (x_1^2, x_2^2, x_1x_2 + x_2x_1).$$

The embedding  $\Lambda^2(\mathbb{C}^3) \times \mathbb{C}^3 \hookrightarrow \mathcal{A} \times \mathcal{A}$  is given by

$$a_1 \mapsto -x_1, \quad a_2 \mapsto -x_2, \quad a_3 \mapsto 1, \quad b_1 \mapsto 1, \quad b_2 \mapsto -x_2, \quad b_3 \mapsto x_1.$$

Then given  $A = \sum_{i=1}^3 u_i a_i \in \Lambda^2(\mathbb{C}^3)$  and  $x = \sum_{i=1}^3 v_i b_i \in \mathbb{C}^3$ , their images  $\widehat{A}, \widehat{x} \in \mathcal{A}$  are given by

$$\widehat{A} = -u_1 x_1 - u_2 x_2 + u_3, \quad \widehat{x} = v_1 - v_2 x_2 + v_3 x_1,$$

and their product is given by

$$\widehat{A} \cdot \widehat{x} = (-u_1 v_1 + u_3 v_3) \cdot x_1 + (u_1 v_2 + u_2 v_3) \cdot x_1 x_2 + (-u_2 v_1 - u_3 v_2) \cdot x_2 + (u_3 v_1) 1 \in \mathcal{A}.$$

Hence  $\mathcal{A}$  realizes  $3 \times 3$  skew-symmetric matrix-vector product. We observe that  $\mathcal{A}$  may be regarded as the cohomology ring of a torus, i.e., an exterior algebra of a two-dimensional vector space.

We now discuss the general case of  $n \times n$  skew-symmetric matrix-vector product  $\beta_\Lambda : \Lambda^2(\mathbb{C}^n) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ . We construct an algebra that realizes  $\beta_\Lambda$  starting with the inclusion of vector spaces

$$\Lambda^2(\mathbb{C}^n) \hookrightarrow (\mathbb{C}[x]/(x^n + 1)) \oplus W, \quad (15)$$

where  $W$  is a linear subspace of  $\mathbb{C}^{n \times n}$  matrices satisfying the following conditions

- (i) entries in the first row are all zeros;
- (ii) diagonal entries are all zeros;
- (iii) entries in the first column satisfy the relation  $a_{i,1} + a_{n+2-i,1} = 0$  for  $i = 2, \dots, n$ .

Given  $A \in \Lambda^2(\mathbb{C}^n)$ , the embedding is given by the decomposition

$$A = \begin{bmatrix} 0 & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ -a_{1,2} & 0 & \cdots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{1,n-1} & -a_{2,n-1} & \cdots & 0 & a_{n-1,n} \\ -a_{1,n} & -a_{2,n} & \cdots & -a_{n-1,n} & 0 \end{bmatrix} = A_c + A_w,$$

where

$$A_c = \begin{bmatrix} 0 & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ -a_{1,n} & 0 & \cdots & a_{1,n-2} & a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{1,3} & -a_{1,4} & \cdots & 0 & a_{1,2} \\ -a_{1,2} & -a_{1,3} & \cdots & -a_{1,n} & 0 \end{bmatrix} \in \text{Circ}_{n,-1}(\mathbb{C})$$

is a skew-circulant matrix and

$$A_w = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ -a_{1,2} + a_{1,n} & 0 & \cdots & a_{2,n-1} - a_{1,n-2} & a_{2,n} - a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1,3} - a_{1,n-1} & a_{1,4} - a_{2,n-1} & \cdots & 0 & a_{n-1,n} - a_{1,2} \\ a_{1,2} - a_{1,n} & a_{1,3} - a_{2,n} & \cdots & a_{1,n} - a_{n-1,n} & 0 \end{bmatrix} \in W.$$

Note that  $A_c$ , being skew-circulant, may be regarded as an element of  $\mathbb{C}[x]/(x^n + 1)$  as we have discussed in Section 8 and we obtain the embedding in (15).

Since we also have

$$\mathbb{C}^n \simeq \mathbb{C}[x]/(x^n + 1),$$

the bilinear map  $\beta_\wedge : \Lambda^2(\mathbb{C}^n) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$  may be realized as follows.

$$\begin{array}{ccc} \Lambda^2(\mathbb{C}^n) \otimes \mathbb{C}^n & \xrightarrow{j} & (\mathbb{C}[x]/(x^n + 1) \oplus W) \otimes \mathbb{C}^n \\ \text{id} \downarrow & & \downarrow \simeq \\ \Lambda^2(\mathbb{C}^n) \otimes \mathbb{C}^n & \xrightarrow{j} & (\mathbb{C}[x]/(x^n + 1) \otimes \mathbb{C}^n) \oplus (W \otimes \mathbb{C}^n) \\ \text{id} \downarrow & & \downarrow \simeq \\ \Lambda^2(\mathbb{C}^n) \otimes \mathbb{C}^n & \xrightarrow{j} & (\mathbb{C}[x]/(x^n + 1) \otimes \mathbb{C}[x]/(x^n + 1)) \oplus (W \otimes \mathbb{C}^n) \\ \beta_\wedge \downarrow & & \downarrow m \\ \mathbb{C}^n & \xleftarrow{\text{proj}} & \mathbb{C}[x]/(x^n + 1) \simeq \mathbb{C}^n \end{array}$$

We have identified the skew-circulant matrix vector product with the skew-circulant matrix-matrix product (cf. Section 8), i.e., the multiplication

$$\mathbb{C}[x]/(x^n + 1) \times \mathbb{C}^n \rightarrow \mathbb{C}^n$$

is identified with the multiplication

$$\mathbb{C}[x]/(x^n + 1) \times \mathbb{C}[x]/(x^n + 1) \rightarrow \mathbb{C}[x]/(x^n + 1).$$

This realization is another instance of the commutative diagram (6). To put all these in concrete terms, given  $A \in \Lambda^2(\mathbb{C}^n)$  and  $x \in \mathbb{C}^n$ , we compute the matrix-vector product via

$$Ax = (\text{first row of } A_c \text{ Circ}(x)) + A_w x.$$

**Theorem 15.3.** *The rank of the structure tensor of skew-symmetric matrix-vector product is bounded above by  $n^2 - n - \lceil (n-1)/2 \rceil + 1$ .*

*Proof.* The first factor of the realization is the multiplication in the algebra

$$\mathbb{C}[x]/(x^n + 1) \times \mathbb{C}[x]/(x^n + 1) \rightarrow \mathbb{C}[x]/(x^n + 1).$$

By Theorem 4.5 we see that the rank of the structure tensor of the algebra  $\mathbb{C}[x]/(x^n + 1)$  is  $n$ . The second factor of the realization is a bilinear map

$$W \times \mathbb{C}^n \rightarrow \mathbb{C}^n.$$

A matrix-vector product with a matrix in  $W$  costs  $n^2 - (2n - 1) - \lceil (n-1)/2 \rceil$  multiplications — there are  $2n - 1$  zeros by (i) and (ii) and we invoke Proposition 6.1; moreover there are  $\lceil (n-1)/2 \rceil$  identical terms by (iii). Therefore this realization gives an upper bound of

$$n^2 - n - \lceil (n-1)/2 \rceil + 1. \quad \square$$

This upper bound is  $n + \lceil (n-1)/2 \rceil - 1$  multiplications fewer than the usual matrix-vector product. In particular, for  $n = 3$  we obtain the upper bound in Proposition 15.2.

## 16. COMMUTATOR

Our study of the bilinear complexity of commutators in this section covers only the case of  $2 \times 2$  matrices. We do not yet know how to extend them to  $n \times n$  matrices when  $n > 2$ .

We consider the bilinear map  $[\cdot, \cdot] : \mathbb{C}^{2 \times 2} \times \mathbb{C}^{2 \times 2} \rightarrow \mathbb{C}^{2 \times 2}$  defined by  $[A, X] = AX - XA$ . We will write

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad X = \begin{bmatrix} x & y \\ z & w \end{bmatrix},$$

and therefore

$$[A, X] = \begin{bmatrix} bz - cy & (a-d)y - b(x-w) \\ -(a-d)z + c(x-w) & -(bz - cy) \end{bmatrix}.$$

Hence the rank of  $\mu_{[\cdot, \cdot]} \in (\mathbb{C}^{2 \times 2})^* \otimes (\mathbb{C}^{2 \times 2})^* \otimes \mathbb{C}^{2 \times 2}$ , the structure tensor of  $[\cdot, \cdot]$ , is at most six.

Now consider the matrix-vector product between matrices and vectors of the following forms

$$\begin{bmatrix} 0 & -c & b \\ -b & a-d & 0 \\ c & 0 & -a-d \end{bmatrix} \in \mathbb{C}^{3 \times 3}, \quad \begin{bmatrix} x-w \\ y \\ z \end{bmatrix} \in \mathbb{C}^3.$$

Notice that

$$\begin{bmatrix} 0 & -c & b \\ -b & a-d & 0 \\ c & 0 & -(a-d) \end{bmatrix} \begin{bmatrix} x-w \\ y \\ z \end{bmatrix} = \begin{bmatrix} bz - cy \\ (a-d)y - b(x-w) \\ -(a-d)z + c(x-w) \end{bmatrix},$$

So the rank and border rank of the structure tensor  $\mu_{[\cdot, \cdot]}$  of  $[\cdot, \cdot]$  is the same as the rank of the following bilinear operation

$$\beta : \mathbb{C}^3 \times \mathbb{C}^3 \rightarrow \mathbb{C}^3, \quad \left( \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}, \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \right) \mapsto \begin{bmatrix} s_1 t_2 + s_2 t_3 \\ -s_2 t_1 + s_3 t_2 \\ -s_1 t_1 - s_3 t_3 \end{bmatrix}, \quad (16)$$

where

$$s_1 = -c, \quad s_2 = b, \quad s_3 = a-d, \quad t_1 = x-w, \quad t_2 = y, \quad t_3 = z.$$

We will need to distinguish the three copies of  $\mathbb{C}^3$  in (16), so for clarity let us denote them by  $U$ ,  $V$ , and  $W$  respectively, i.e.,

$$\beta : U \times V \rightarrow W.$$

Let  $\{u_1, u_2, u_3\}$ ,  $\{v_1, v_2, v_3\}$ ,  $\{w_1, w_2, w_3\}$  be the standard bases of  $U, V, W$ . Then the structure tensor  $\mu_\beta$  of  $\beta$  may be decomposed as

$$\mu_\beta = (u_1 \otimes v_2 + u_2 \otimes v_3) \otimes w_1 + (-u_2 \otimes v_1 + u_3 \otimes v_2) \otimes w_2 + (-u_1 \otimes v_1 - u_3 \otimes v_3) \otimes w_3$$

and we may apply Theorem 15.1 to obtain the following.

**Corollary 16.1.** *The rank and border rank of the commutator for  $2 \times 2$  matrices are given by*

$$\text{rank}(\mu_{[\cdot, \cdot]}) = 5 \quad \text{or} \quad 6$$

and

$$\text{brank}(\mu_{[\cdot, \cdot]}) = 5$$

respectively.

In other words, for  $A, X \in \mathbb{C}^{2 \times 2}$ , computing  $AX$  requires at least seven multiplications (e.g. Strassen's algorithm) whereas computing  $[A, X] = AX - XA$  requires at most six multiplications. We suspect that this is always the case, i.e., computing commutator is always faster than computing matrix multiplication for  $n \times n$  matrices.

We now construct an algebra  $\mathcal{A}$  that realizes  $\beta$  and therefore  $[\cdot, \cdot]$ . Let

$$\mathcal{A} = \mathbb{C}\langle x_1, x_2 \rangle / (x_1^2, x_2^2, x_1 x_2 + x_2 x_1)$$

and consider the embedding  $U \otimes V \rightarrow \mathcal{A} \otimes \mathcal{A}$  induced by

$$u_1 \mapsto x_1, \quad u_2 \mapsto x_2, \quad u_3 \mapsto 1, \quad v_1 \mapsto -1, \quad v_2 \mapsto x_2, \quad v_3 \mapsto -x_1.$$

Given  $s = \sum_{i=1}^3 s_i u_i \in U$  and  $t = \sum_{i=1}^3 t_i v_i \in V$ , the images  $\widehat{s}$  and  $\widehat{t}$  in  $\mathcal{A}$  are

$$\widehat{s} = -s_1 x_1 - s_2 x_2 + s_3 1, \quad \widehat{t} = -t_1 1 + t_2 x_2 - t_3 x_1$$

respectively. Their product is

$$\begin{aligned} \widehat{s} \cdot \widehat{t} &= (s_1 x_1 + s_2 x_2 + s_3 1)(-t_1 1 + t_2 x_2 - t_3 x_1) \\ &= (-s_1 t_1 - s_3 t_3) x_1 + (s_1 t_2 + s_2 t_3) x_1 x_2 + (-s_2 t_1 + s_3 t_2) x_2 + (-s_3 t_1) 1, \end{aligned}$$

i.e.,  $\mathcal{A}$  realizes the bilinear map  $[\cdot, \cdot]$ . Observe that  $\mathcal{A}$  is the same algebra that we used to realize the  $3 \times 3$  skew-symmetric matrix-vector product in Section 15.

## 17. SIMULTANEOUS MATRIX MULTIPLICATION

We round out our list of bilinear operations with two examples of *simultaneous matrix product*.

**Proposition 17.1.** *The following two matrix-matrix products:*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} g & h \\ e & f \end{bmatrix} \quad (17)$$

can be computed simultaneously with eight multiplications.

*Proof.* Let  $D_4 = \langle x, y \mid x^4 = y^2 = 1, \quad yxy = x^{-1} \rangle$  be the dihedral group of order eight. The multiplication of  $2 \times 2$  matrices is realized by the subsets

$$H_1 = \langle y \rangle = \{y, 1\}, \quad H_2 = \langle x^2 y \rangle = \{x^2 y, 1\}, \quad S_3 = \{x^{-1} y, 1\}.$$

Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}. \quad (18)$$

Then  $A, B$  correspond to  $\widehat{A}, \widehat{B} \in \mathbb{C}[D_4]$  where

$$\begin{aligned} \widehat{A} &= a \cdot (y^{-1} x^2 y) + b \cdot (y^{-1}) + c \cdot (x^2 y) + d \cdot (1) \\ &= a \cdot (x^2) + b \cdot (y) + c \cdot (x^2 y) + d \cdot (1), \\ \widehat{B} &= e \cdot ((x^2 y)^{-1} x^{-1} y) + f \cdot ((x^2 y)^{-1}) + g \cdot (x^{-1} y) + h \cdot (1) \\ &= e \cdot (x^3) + f \cdot (x^2 y) + g \cdot (x^3 y) + h \cdot (1). \end{aligned}$$

We compute the product  $\widehat{A} \cdot \widehat{B}$  in  $\mathbb{C}[D_4]$ ,

$$\begin{aligned} \widehat{A} \cdot \widehat{B} &= (ae + bg) \cdot (x) + (af + bh) \cdot (y) + (ce + dg) \cdot (x^3 y) + (cf + dh) \cdot (1) \\ &\quad + (ag + be) \cdot (xy) + (ah + bf) \cdot (x^2) + (cg + de) \cdot (x^3) + (ch + df) \cdot (x^2 y) \end{aligned}$$

and observe that the first four terms and last four terms are precisely the entries of

$$M_1 := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad \text{and} \quad M_2 := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} g & h \\ e & f \end{bmatrix}$$

respectively. In other words, we can calculate  $M_1$  and  $M_2$  simultaneously by calculating  $\widehat{A} \cdot \widehat{B}$ . On the other hand,  $D_4$  has four irreducible representations of dimension one and one of dimension two:

- (i) trivial:  $(1, x, x^2, x^3, y, xy, x^2 y, x^3 y) \xrightarrow{\rho} (1, 1, 1, 1, 1, 1, 1, 1)$
- (ii) sign type 1:  $(1, x, x^2, x^3, y, xy, x^2 y, x^3 y) \xrightarrow{\rho} (1, 1, 1, 1, -1, -1, -1, -1)$
- (iii) sign type 2:  $(1, x, x^2, x^3, y, xy, x^2 y, x^3 y) \xrightarrow{\rho} (1, -1, 1, -1, 1, -1, 1, -1)$
- (iv) sign type 3:  $(1, x, x^2, x^3, y, xy, x^2 y, x^3 y) \xrightarrow{\rho} (1, -1, 1, -1, -1, 1, -1, 1)$

(v) two-dimensional:

$$\begin{aligned} 1 &\mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & x &\mapsto \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, & x^2 &\mapsto \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, & x^3 &\mapsto \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \\ y &\mapsto \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, & xy &\mapsto \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & x^2y &\mapsto \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, & x^3y &\mapsto \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}. \end{aligned}$$

By Wedderburn Theorem we have

$$\mathbb{C}[D_4] \simeq \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}^2 \otimes \mathbb{C}^2,$$

where the first four  $\mathbb{C}$ 's correspond to the four 1-dimensional representations and the  $\mathbb{C}^2 \otimes \mathbb{C}^2$  corresponds to the 2-dimensional representation. Under this isomorphism, we may identify  $\widehat{A}$  and  $\widehat{B}$  as  $6 \times 6$  block diagonal matrices,

$$\widehat{A} = \begin{bmatrix} a+b+c+d & 0 & 0 & 0 & 0 & 0 \\ 0 & a-b-c+d & 0 & 0 & 0 & 0 \\ 0 & 0 & a+b+c+d & 0 & 0 & 0 \\ 0 & 0 & 0 & a-b-c+d & 0 & 0 \\ 0 & 0 & 0 & 0 & -a+b-c+d & 0 \\ 0 & 0 & 0 & 0 & 0 & -a-b+c+d \end{bmatrix}$$

and

$$\widehat{B} = \begin{bmatrix} e+f+g+h & 0 & 0 & 0 & 0 & 0 \\ 0 & e-f-g+h & 0 & 0 & 0 & 0 \\ 0 & 0 & -e+f-g+h & 0 & 0 & 0 \\ 0 & 0 & 0 & -e-f+g+h & 0 & 0 \\ 0 & 0 & 0 & 0 & -f+h & e-g \\ 0 & 0 & 0 & 0 & -e-g & f+h \end{bmatrix}.$$

Hence the computation of  $\widehat{A} \cdot \widehat{B}$  costs eight multiplications.  $\square$

**Corollary 17.2.** *Suppose*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} g & h \\ e & f \end{bmatrix} = 0.$$

*Then the product*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

*can be computed with four multiplications.*

*Proof.* If the given condition holds, one may obtain the required product from the first four diagonal entries of  $\widehat{A} \cdot \widehat{B}$ , which costs four multiplications.  $\square$

We restate Proposition 17.1 in terms of the structure tensor  $\mu_f$  of the bilinear map

$$\beta_f : \mathbb{C}^{2 \times 2} \times \mathbb{C}^{2 \times 2} \rightarrow \mathbb{C}^{2 \times 2} \oplus \mathbb{C}^{2 \times 2}, \quad (A, B) \mapsto (AB, AB^f),$$

where  $B^f$  denotes the operation of switching the first and second row of  $B$ . Note that

$$\mu_f \in (\mathbb{C}^{2 \times 2})^* \otimes (\mathbb{C}^{2 \times 2})^* \otimes (\mathbb{C}^{2 \times 2} \oplus \mathbb{C}^{2 \times 2}) \simeq \mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^8.$$

**Proposition 17.3.** *The rank and border rank of the structure tensor for the simultaneous matrix-product in (17) are given by*

$$\text{rank}(\mu_f) = \text{brank}(\mu_f) = 8.$$

*Proof.* It is easy to verify that  $\text{span}(\mu_f(\mathbb{C}^{2 \times 2} \otimes \mathbb{C}^{2 \times 2})) = \mathbb{C}^{2 \times 2}$ . Hence the required result follows from Propositions 3.2, 3.4, and 17.1.  $\square$

**Corollary 17.4.** *Consider the matrices*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{C}^{2 \times 2}, \quad B = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \\ f_1 & f_2 & \cdots & f_n \end{bmatrix} \in \mathbb{C}^{2 \times n}, \quad B^f = \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ e_1 & e_2 & \cdots & e_n \end{bmatrix} \in \mathbb{C}^{2 \times n}, \quad (19)$$

where  $n$  is any positive integer. Then  $AB$  and  $AB^f$  can be computed simultaneously with  $8n$  multiplications.

*Proof.* We may realize the bilinear map

$$\mathbb{C}^{2 \times 2} \times \mathbb{C}^{2 \times n} \rightarrow \mathbb{C}^{2 \times n} \oplus \mathbb{C}^{2 \times n}, \quad (A, B) \mapsto (AB, AB^f)$$

by the algebra  $\mathbb{C}[D_4] \times \cdots \times \mathbb{C}[D_4]$  ( $n$  copies).  $\square$

Suppose we are instead interested in computing

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} h & g \\ e & f \end{bmatrix}$$

simultaneously. We start by realizing  $2 \times 2$  matrix product by the algebra  $\mathbb{C}[x]/(x^8 - 1)$ . Let  $A, B \in \mathbb{C}^{2 \times 2}$  be as in (18). Consider the embedding

$$j : \mathbb{C}^{2 \times 2} \otimes \mathbb{C}^{2 \times 2} \rightarrow \mathbb{C}[x]/(x^8 - 1) \otimes \mathbb{C}[x]/(x^8 - 1), \\ (A, B) \mapsto (ax^3 + cx^2 + bx + d, gx^6 + ex^4 + hx^2 + f),$$

and the projection

$$\text{proj} : \mathbb{C}[x]/(x^8 - 1) \rightarrow \mathbb{C}^{2 \times 2}, \quad \sum_{i=1}^7 u_i x^i \mapsto \begin{bmatrix} u_7 & u_3 \\ u_6 & u_2 \end{bmatrix}.$$

We may verify that for these choices, the diagram in (5) commutes. The product

$$(ax^3 + cx^2 + bx + d)(gx^6 + ex^4 + hx^2 + f)$$

in  $\mathbb{C}[x]/(x^8 - 1)$  gives us the following counterpart of Proposition 17.1.

**Proposition 17.5.** *The following two matrix-matrix products:*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} h & g \\ e & f \end{bmatrix} \quad (20)$$

can be computed simultaneously with eight multiplications.

Again, we may restate Proposition 17.5 in terms of the structure tensor  $\mu_{\mathbf{g}}$  of the bilinear map

$$\beta_{\mathbf{g}} : \mathbb{C}^{2 \times 2} \times \mathbb{C}^{2 \times 2} \rightarrow \mathbb{C}^{2 \times 2} \oplus \mathbb{C}^{2 \times 2}, \quad (A, B) \mapsto (AB, AB^{\mathbf{g}})$$

where  $B^{\mathbf{g}}$  is the matrix obtained from  $B$  by switching the first row and the second row and then switching the first and the second entry in the first row. The following analogue of Proposition 17.3 follows from Propositions 3.2, 3.4, and 17.5.

**Proposition 17.6.** *The rank and border rank of the structure tensor for the simultaneous matrix-matrix product in (20) are given by*

$$\text{rank}(\mu_{\mathbf{g}}) = \text{brank}(\mu_{\mathbf{g}}) = 8.$$

We also have the following analogue of Corollary 17.4.

**Corollary 17.7.** *Let  $A \in \mathbb{C}^{2 \times 2}$  and  $B \in \mathbb{C}^{2 \times n}$  be as in (19). Let  $B^{\mathbf{g}} \in \mathbb{C}^{2 \times n}$  be the matrix obtained from  $B$  by switching the first and second row followed by switching  $2i$ th and  $(2i - 1)$ th entry in the first row for  $i = 1, 2, \dots, \lfloor n/2 \rfloor$ . Then  $AB$  and  $AB^{\mathbf{g}}$  can be computed simultaneously with  $8n$  multiplications.*

*Proof.* We may realize the bilinear map

$$\mathbb{C}^{2 \times 2} \times \mathbb{C}^{2 \times n} \rightarrow \mathbb{C}^{2 \times n} \oplus \mathbb{C}^{2 \times n}, \quad (A, B) \mapsto (AB, AB^g)$$

by the algebra  $\mathbb{C}[x_1, \dots, x_n]/(x_i^8 - 1 \mid i = 1, \dots, n)$ .  $\square$

## 18. CONCLUSION

The Strassen tensor rank approach gives us a simple way for quantifying bilinear complexity whereas the (generalized) Cohn–Umans approach gives us a constructive way that allows for the rich properties of various algebras to be used in analyzing bilinear complexity. The two methods can be applied hand-in-hand to systematically discover algorithms of optimal bilinear complexity.

**Acknowledgment.** We thank Henry Cohn for very helpful discussions that initiated this work. We are also grateful to Andrew Chien, Nikos Pitsianis, and Xiaobai Sun for answering our questions about energy costs and circuit complexity of various integer and floating point operations; to Mike Stein for suggesting that we examine BTB matrices; and to Chris Umans for prompting the noncommutative construction at the end of Section 5.

LHL and KY are partially supported by AFOSR FA9550-13-1-0133, DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, and DMS 1057064. In addition, KY’s work is also partially supported by NSF CCF 1017760.

## REFERENCES

- [1] D. Bini and M. Capovani, “Tensor rank and border rank of band Toeplitz matrices,” *SIAM J. Comput.*, **16** (1987), no. 2, pp. 252–258.
- [2] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.
- [3] J.-L. Brylinski, “Algebraic measures of entanglement,” pp. 3–23, G. Chen and R. K. Brylinski (Eds), *Mathematics of Quantum Computation*, CRC, Boca Raton, FL, 2002.
- [4] J. Buczyński and J. M. Landsberg, “Ranks of tensors and a generalization of secant varieties,” *Linear Algebra Appl.*, **438** (2013), no. 2, pp. 668–689.
- [5] P. Bürgisser, M. Clausen, and M. A. Shokrollahi, *Algebraic Complexity Theory*, Grundlehren der Mathematischen Wissenschaften, **315**, Springer-Verlag, Berlin, 1997.
- [6] R. H.-F. Chan and X.-Q. Jin, *An Introduction to Iterative Toeplitz Solvers*, Fundamentals of Algorithms, **5**, SIAM, Philadelphia, PA, 2007.
- [7] H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans, “Group-theoretic algorithms for matrix multiplication,” *Proc. IEEE Symp. Found. Comput. Sci. (FOCS)*, **46** (2005), pp. 379–388.
- [8] H. Cohn and C. Umans, “A group-theoretic approach to fast matrix multiplication,” *Proc. IEEE Symp. Found. Comput. Sci. (FOCS)*, **44** (2003), pp. 438–449.
- [9] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” *Math. Comp.*, **19** (1965), no. 90, pp. 297–301.
- [10] D. Coppersmith and S. Winograd, “Matrix multiplication via arithmetic progressions,” *J. Symbolic Comput.*, **9** (1990), no. 3, pp. 251–280.
- [11] V. De Silva and L.-H. Lim, “Tensor rank and the ill-posedness of the best low-rank approximation problem,” *SIAM J. Matrix Anal. Appl.*, **30** (2008), no. 3, pp. 1084–1127.
- [12] G. Golub and C. Van Loan, *Matrix Computations*, 4th Ed., John Hopkins University Press, Baltimore, MD, 2013.
- [13] N. J. Higham, *Functions of Matrices*, SIAM, Philadelphia, PA, 2008.
- [14] *Intel 64 and IA-32 Architectures Optimization Reference Manual*, September 2015, <http://www.intel.com/content/dam/www/public/us/en/documents/manuals/64-ia-32-architectures-optimization-manual>
- [15] T. Kailath and J. Chun, “Generalized displacement structure for block-Toeplitz, Toeplitz-block, and Toeplitz-derived matrices,” *SIAM J. Matrix Anal. Appl.*, **15** (1994), no. 1, pp. 114–128.
- [16] T. Kailath and A. H. Sayed, *Fast Reliable Algorithms for Matrices with Structure*, SIAM, Philadelphia, PA, 1999.
- [17] D. E. Knuth, *The Art of Computer Programming*, Volume 2: Seminumerical algorithms, 3rd Ed., Addison–Wesley, Reading, MA, 1998.
- [18] V. K. Kodavalla, “IP gate count estimation methodology during micro-architecture phase,” *IP Based Electronic System Conference and Exhibition (IP-SOC)*, Grenoble, France, December 2007, <http://www.design-reuse.com/articles/19171/ip-gate-count-estimation-micro-architecture-phase.html>



- [19] J. M. Landsberg, *Tensors: Geometry and Applications*, Graduate Studies in Mathematics, **128**, AMS, Providence, RI, 2012.
- [20] F. Le Gall, “Powers of tensors and fast matrix multiplication,” *Proc. Internat. Symp. Symbolic Algebr. Comput. (ISSAC)*, **39** (2014), pp. 296–303.
- [21] L.-H. Lim, “Tensors and hypermatrices,” in: L. Hogben (Ed.), *Handbook of Linear Algebra*, 2nd Ed., CRC Press, Boca Raton, FL, 2013.
- [22] J. C. McConnell and J. C. Robson, *Noncommutative Noetherian Rings*, Rev. Ed., Graduate Studies in Mathematics, **30**, AMS, Providence, RI, 2001.
- [23] M. K. Ng, *Iterative Methods for Toeplitz Systems*, Oxford University Press, New York, NY, 2004.
- [24] G. Ottaviani, “Symplectic bundles on the plane, secant varieties and Lüroth quartics revisited,” *Quad. Mat.*, **21** (2007), pp. 315–352.
- [25] G. Strang and S. MacNamara, “Functions of difference matrices are Toeplitz plus Hankel,” *SIAM Rev.*, **56** (2014), no. 3, pp. 525–546.
- [26] V. Strassen, “Gaussian elimination is not optimal,” *Numer. Math.*, **13** (1969), pp. 354–356.
- [27] V. Strassen, “Rank and optimal computation of generic tensors,” *Linear Algebra Appl.*, **52/53** (1983), pp. 645–685.
- [28] V. Strassen, “Relative bilinear complexity and matrix multiplication,” *J. Reine Angew. Math.*, **375/376** (1987), pp. 406–443.
- [29] V. Strassen, “Vermeidung von Divisionen,” *J. Reine Angew. Math.*, **264** (1973), pp. 184–202.
- [30] C. F. Van Loan, “The ubiquitous Kronecker product,” *J. Comput. Appl. Math.*, **123** (2000), no. 1–2, pp. 85–100.
- [31] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*, SIAM, Philadelphia, PA, 2007.
- [32] V. V. Williams, “Multiplying matrices in  $O(n^{2.373})$  time,” *preprint*, <http://theory.stanford.edu/~virgi/matrixmult-f.pdf>, Stanford, CA, 2014.
- [33] S. Winograd, “Some bilinear forms whose multiplicative complexity depends on the field of constants,” *Math. Systems Theory*, **10** (1976/77), no. 2, pp. 169–180.
- [34] K. Ye and L.-H. Lim, “Every matrix is a product of Toeplitz matrices,” *Found. Comput. Math.*, **16** (2016), no. 1, to appear.

DEPARTMENT OF MATHEMATICS AND DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.

*E-mail address:* [kye@math.uchicago.edu](mailto:kye@math.uchicago.edu)

COMPUTATIONAL AND APPLIED MATHEMATICS INITIATIVE, DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.

*E-mail address, corresponding author:* [lekheng@galton.uchicago.edu](mailto:lekheng@galton.uchicago.edu)