GEOMETRIC CONFIGURATIONS OF ALGORITHMS FOR REDUCED $M \times 2$ AND 2×2 MATRIX MULTIPLICATION

A Thesis

by

BINGJIN LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Joseph M. Landsberg
Committee Members,	Igor Zelenko
	Jennifer L. Welch
Head of Department,	Emil J. Straube

May 2016

Major Subject: Mathematics

ABSTRACT

Matrix multiplication is commonly used in scientific computation. Given matrices $A = (a_{ij})$ of size $l \times m$ and $B = (b_{ij})$ of size $m \times n$, the standard way to compute the product $C := AB$ is computing $c_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj}$. In this case, lmn multiplications and $ln(m-1)$ additions are used. In 1969, V. Strassen found a surprising algorithm to multiply 2×2 matrices using 7 multiplications instead of 8 in the standard algorithm. In this way, $n \times n$ matrix multiplication can be computed using $O(n^{log_2^7})$ scalar multiplication operations. If n is large, the Strassen algorithm is much more efficient than the standard algorithm. After Strassen's algorithm, numerous efforts were made to reduce the complexity for $n \times n$ matrix multiplication. By 1986, the bound was reduced to $O(n^{2.38})$ by Coppersmith and Winograd. However this is an asymptotic result rather than an implementable algorithm. The complexity has not been significantly improved for 30 years.

Matrix multiplication is a tensor and one way to measure the complexity is using its tensor rank. Any tensor can be written as finite sum of rank one tensors and the rank for a tensor is the least number of rank-one tensors needed in the sum. A theorem due to Strassen shows the tensor rank is a good measurement for the complexity. One Bini's theorem demonstrates that the border rank of the matrix multiplication tensor $M_{<n>}$ is a complexity measurement for matrix multiplication . Even though the problem may sound simple, the border ranks of small matrix multiplication tensors are still unknown. Suppose one wants to compute the border rank of the tensor for the matrix multiplication of size $m \times 2$ and 2×2 denoted by $\underline{R}(M_{<m,2,2>})$. $\underline{R}(M_{<m,2,2>})$ is closely related to the border rank of reduced matrix multiplication tensor $T_{BCLRS,m}$, where one entry is set equal to zero. For small m

like 2 and 3, there are good geometric configurations in the border rank algorithms for the tensor $T_{BCLRS,m}$. My project is to understand the geometry of the good existing algorithms in the cases $m = 2, 3$. In the configuration of case $m = 2$, the limit 5-plane in the Grassmannian plane in the algorithm intersects with the Segre variety in three special lines. For the case $m = 3$, the intersection of the limiting 8-plane and the Segre variety consists of the union of a family of lines passing through a plane conic and a special sub-Segre variety. I also try to find analogous algorithms to the $m = 2$ case or disprove the existence of such algorithms.

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Professor J.M. Landsberg at Texas A&M University. He helped me whenever I had some problems in the research or in my writing.

I would like to thank my friend Yonghui Guan, a PhD student at Texas A&M University. He would explain the concepts which I was confused about.

Finally, I would like to thank my roommate Yanhe Huang and my parents. They supported me whenever I had a hard time in my master study.

TABLE OF CONTENTS

Page

1. INTRODUCTION

1.1 Tensor products of vector spaces

1.1.1 Dual space

For a vector space V of dimension n, the *dual vector space* of V is defined as all linear maps from V to $\mathbb{C}$. It is a vector space over $\mathbb{C}$ under the addition and scalar multiplication of linear maps.

$$V^* := \{f : V \to \mathbb{C} | f \textit{ is linear}\}$$

Given a basis of V denoted by $\{v_1, v_2, \ldots, v_n\}$, there is a natural *dual basis* for V^*–$\{v_i^*\}$. v_i^* is defined on the basis of V by $v_i^*(v_j) = \delta_{i,j}$, then extends to V linearly. V and V^* is of the same dimension. We can identify elements in V as column vectors after fixing a basis in V, and identify elements in V^* with row vectors under its dual basis. Then for the given basis, there is an isomorphism between V and V^* simply by transposing the column vector to the row one, furthermore for $f \in V^*, v \in V$, $f(v)$ is just the row-column matrix multiplication.

$$f(v) = \begin{pmatrix} a_1, \ldots, a_n \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i$$

with $f = \sum_{i=1}^{n} a_i v_i^*$ and $v = \sum_{i=1}^{n} b_i v_i$.

When V is a finite dimension vector space, there is a canonical vector space isomorphism Φ between V and $(V^*)^*$ defined by

$$\Phi(v)(f) = f(v)$$

with $v \in V$ and $f \in V^*$. Easy to check Φ is a one-one linear map and dim$V$= dim$(V^*)^*$, thus Φ is an isomorphism.

1.1.2 Definitions for tensor products

Next we define the tensor product of two vector spaces A and B in several ways, where dim A= n and dim B=m.

1. $A \otimes_1 B := \{f : A^* \to B | \ f \textit{ is linear}\}$

2. $A \otimes_2 B :=$ linear span of $\{a \otimes_2 b; a \in A, b \in B | (a_1 + a_2) \otimes_2 b = a_1 \otimes_2 b + a_2 \otimes_2 b,\ a \otimes_2 (b_1 + b_2) = a \otimes_2 b_1 + a \otimes_2 b_2,\ ka \otimes_2 b = a \otimes_2 kb; a_1, a_2, a \in A; b_1, b_2, b \in B, k \in \mathbb{C}\}$

3. $A \otimes_3 B := \{T : A^* \times B^* \to \mathbb{C} | \, T \textit{ is bilinear}\}$

Given a basis $\{a_i\}_{i=1}^n$ of A and a basis $\{b_j\}_{j=1}^m$ of B, it is easy to check:

1. $A \otimes_1 B$ is a mn dimensional vector space with basis $a_i \otimes_1 b_j$. Here $a_i \otimes_1 b_j(a_k^*) = \delta_{i,k} b_j$ then extends to A^* linearly.

2. $A \otimes_2 B$ is a vector space with basis $a_i \otimes_2 b_j$, it is of dimension mn.

3. $A \otimes_3 B$ is a vector space with basis $a_i \otimes_3 b_j$. And $a_i \otimes_3 b_j(a_k^*, b_l^*) = \delta_{i,k}\delta_{j,l}$, then extends linearly.

From the argument above, we can see these three definitions are equivalent by identifying their bases and checking the linear structure. So we can just write the tensor product using $A \otimes B$.

Remark. There is a canonical isomorphism Φ_{13} between $A \otimes_1 B$ and $A \otimes_3 B$ defined by

$$\Phi_{13}(f)(a^*, b^*) = b^*(f(a^*))$$

Easy to check that Φ_{13} is injective linear map between two vector spaces having the same dimension. Thus it is an isomorphism. There is also a canonical isomorphism Φ_{21} between $A \otimes_2 B$ and $A \otimes_1 B$ defined by

$$\Phi_{21}(a \otimes_2 b)(f) = f(a)b$$

then extends linearly to $A \otimes_2 B$. Also we can show Φ_{21} is injective,

Remark. From the definition 2 and 3, we can see the order of A, B doesn't matter. (ie. $A \otimes B \cong B \otimes A$).

Next we extend the notion of tensor products to n vector spaces. Let $A_1, \ldots, A_n$ be vector spaces over $\mathbb{C}$. Define by induction $A_1 \otimes \cdots \otimes A_n := (A_1 \otimes \cdots A_{n-1}) \otimes A_n$.

Claim. It is the space of n-linear maps $A_1^* \times \cdots \times A_n^* \to \mathbb{C}$. $A_1^* \otimes \cdots \otimes A_n^* \cong (A_1 \otimes \cdots \otimes A_n)^*$. Any element $a_1 \otimes \cdots \otimes a_n \in A_1 \otimes \cdots \otimes A_n$ is defined as $a_1 \otimes \cdots \otimes a_n(\beta^1, \ldots, \beta^n) = a_1(\beta^1) \cdots a_n(\beta^n)$ for any $\beta^1 \in A_1^*, \ldots, \beta^n \in A_n^*$

We can prove the claim by induction. Since any $f \in (A_1 \otimes \cdots \otimes A_{n-1}) \otimes A_n$, f is a map $(A_1 \otimes \cdots \otimes A_{n-1})^* \to A_n$. It is also a map $A_1^* \otimes \cdots \otimes A_{n-1}^* \to A_n$. Use $\bar{f}$ to denote the n-linear map induced by f. $\bar{f}(a^1, \ldots, a^{n-1}, a^n) := a^n(f(a^1 \otimes \cdots \otimes a^{n-1}))$.

1.2 Complexity of tensors

1.2.1 Tensor rank

Definition. $T \in A_1 \otimes \cdots \otimes A_n$ is said to have *rank one* if there exists $a_i \in A_i$ for any $i = 1, \ldots, n$ such that $T = a_1 \otimes \cdots \otimes a_n$.

Definition. $T \in A_1 \otimes \cdots \otimes A_n$, the *rank* of T is defined to be the smallest r such that T can be written as the sum of r rank one tensors. Denoted by $\mathbf{R}(T) = r$.

Remark. For matrix $A \in \mathbb{C}^{l \times m}$, it can be viewed as an element f in $V^* \otimes U$ with $dimU = m$ and $dimV = l$. Claim that $\mathbf{R}(f) =$ rank of A (also rank of f as a linear map). So the rank of a tensor is a natural generalization of rank of a matrix.

To see this, write $A = (A_1, \ldots, A_m)$ with $A_m \in \mathbb{C}^{l \times 1}$, we can assume after reordering $\{A_1, \ldots, A_r\}$ is linearly independent with $r =$ rank of A. Then $\{A_{r+1}, \ldots, A_m\} \subset span\{A_1, \ldots, A_r\}$. So $f = \sum_{i=1}^{m} e^i \otimes A_i$, here e^i is a fixed basis of V^*, and A_i is the element in U with coordinate expression A_i, substitute $A_{r+1}, \ldots, A_m$ with $A_1, \ldots, A_r$. Then $f = \sum_{i=1}^{r} v^i \otimes A_i$ for some $v^i \in V^*$. So $\mathbf{R}(f) \leq r$.

For the other inequality, by the same argument above we know if $f = \sum_{i=1}^{\mathbf{R(f)}} b^i \otimes u_i$, $\{b^i\}$ must be linear independent. Thus it can be extended to a basis for V^*. The rank of f is the dimension of $Image(f)$=span of all u_i. Thus rank of $f \leq \mathbf{R}(f)$.

1.2.2 Border rank

Definition. Let $T \in A_1 \otimes \cdots \otimes A_n$, the *border rank* of T is the least number r such that $T = \lim_{t \to 0} T_1(t) + \cdots + T_r(t)$ with $T_i(t)$ of rank one for any $t \neq 0$ and any $i = 1, \ldots, r$. The *border rank* is denoted by $\underline{\mathbf{R}}(T) = r$.

Remark.

1. In the definition given above, it is equivalent to say T lies in the limiting r-plane of the r-planes spanned by $T_1(t), \ldots, T_r(t)$.

2. Another way to define the border rank of a tensor is via the Zariski topology. Use $\hat{\sigma}_r^0 \subset A_1 \otimes \cdots \otimes A_n$ to denote all tensors with rank at most r. $\hat{\sigma}_r$ is the Zariski closure of $\hat{\sigma}_r^0$, $\underline{\mathbf{R}}(T)$ can also be defined as the smallest r such that $T \in \hat{\sigma}_r$, the two definitions are equivalent since the Euclidean and the Zariski closure of $\hat{\sigma}_r^0$ are the same.(see [6], §2.4.3)

3. $\mathbf{R}(T) \geq \underline{\mathbf{R}}(T)$. Furthermore, if $T \in A_1 \otimes A_2$, $\mathbf{R}(T) = \underline{\mathbf{R}}(T)$, because in this case $\hat{\sigma}_r^0 \subset A_1 \otimes A_2$ is a closed set and can be viewed as matrices with all the $(r+1)$ minors zero.

Example. Let $T = a_1 \otimes b_1 \otimes c_2 + a_1 \otimes b_2 \otimes c_1 + a_2 \otimes b_1 \otimes c_1 \in A_1 \otimes A_2 \otimes A_3$, with $\{a_1, a_2\}$, $\{b_1, b_2\}$ and $\{c_1, c_2\}$ linearly independent.
$\mathbf{R}(T) = 3$.

If not, case 1: $\mathbf{R}(T) = 1$; assume $T = a \otimes b \otimes c$; case 2: $\mathbf{R}(T) = 2$ assume $T = u_1 \otimes v_1 \otimes w_1 + u_2 \otimes v_2 \otimes w_2$, we can assume that $\{u_1, u_2\}$ are linear independent. T can also be viewed as a map $A_1^* \to A_2 \otimes A_3$. In case 2, assume $u_1 = \lambda_1 a_1 + \lambda_2 b_1$ with $\lambda_1 \neq 0$. $T(u_1^*) = v_1 \otimes w_1$, and it is of rank 1. However, $T(u_1^*) = \lambda_1(b_1 \otimes c_2 + b_2 \otimes c_1) + \lambda_2(b_1 \otimes c_1)$, which is of rank 2. Similarly we can say case 1 is not valid either.

But $\underline{\mathbf{R}}(T) = 2$. Since we have $T = \lim_{t \to 0} \frac{1}{t}[(a_1 + ta_2) \otimes (b_1 + tb_2) \otimes (c_1 + tc_2) - a_1 \otimes b_1 \otimes c_1]$, so $\underline{\mathbf{R}}(T) \leq 2$, and the border rank can't be one. Since if it is, T will be of rank one.

1.3 Relation between tensors and matrix multiplication

Consider U, V, W are three vector spaces over $\mathbb{C}$ of dimension l, m, n respectively. $\{u_1, \ldots, u_l\}$ is one basis for U, $\{u^1, \ldots, u^l\}$ is dual basis of U^*. Use column vectors to denote elements in U, while row vectors are for elements in U^*. Similar conventions are for V and W. By definition above, $U^* \otimes V$ is the set of all linear maps from U to V (using canonical isomorphism between U and $(U^*)^*$). $u^i \otimes v_j$ corresponds the map $u^i \otimes v_j(u_k) = \delta_{ij} v_k$ and corresponding matrix is $E_{ji} := (e_j^{(m)})^T e_i^{(l)} \in \mathbb{C}^{m \times l}$. Here $e_i^{(l)} \in \mathbb{C}^{1 \times l}$ and its only nonzero entry is ith entry with value 1, E_{ji} is just the matrix having only nonzero entry at jth row ith column with value 1, here T means the transpose of a matrix.

Every rank one tensor $u^* \otimes v$ is the map $u^* \otimes v(a) = u^*(a)v$, corresponding matrix is the product of a column vector and a row vector, the two vectors are just coordinates of v and u^* respectively.

Claim. There is a canonical isomorphism between $(U^* \otimes V)^*$ and $U \otimes V^*$.

View $u \otimes v^*$ as a linear map by $u \otimes v^*(a^* \otimes b) = a^*(u)v^*(b)$. To see this in basis, $A \in \mathbb{C}^{l\times m}$ ($A \in U \otimes V^*$) and $B \in \mathbb{C}^{m\times l}$ ($B \in V \otimes U^*$), $A(B) = \sum_{i=1}^{l}\sum_{j=1}^{m} a_{ij}b_{ji} = trace(AB)$.

Next we will view matrix multiplication as a tensor.

Let $A \in \mathbb{C}^{l\times m}$ (thus in $V \otimes U^* \cong (U^* \otimes V)^*$) and $B \in \mathbb{C}^{m\times n}$ (in $(U^* \otimes V)^*$), then AB is in $\mathbb{C}^{l\times n} \cong W^* \otimes U$. Matrix multiplication M is a bilinear map

$$M_{<l,m,n>} : (U^* \otimes V)^* \times (V^* \otimes W)^* \to W^* \otimes U$$

which can also be viewed as a trilinear map

$$M_{<l,m,n>} : (U^* \otimes V)^* \times (V^* \otimes W)^* \times (W^* \otimes U)^* \to \mathbb{C}$$

In basis the trilinear map sends (X, Y, Z) to $trace(XYZ)$.

Since $(XY)_{ij} = \sum_{k=1}^{m} x_{ik}y_{kj}$ for $i = 1, \ldots, l$ and $j = 1, \ldots, n$, $M_{<l,m,n>}(X, Y, Z) = XY(Z) = \sum_{i=1}^{l}\sum_{j=1}^{n}(XY)_{ij}Z_{ji} = \sum_{i=1}^{l}\sum_{k-1}^{m}\sum_{j=1}^{n} x_{ik}y_{kj}z_{ji} = trace(XYZ)$ (by the last claim). By the definition, we also have $M_{<l,m,n>} \in (U^* \otimes V) \otimes (V^* \otimes W) \otimes (W^* \otimes U) \cong (U \otimes U^*) \otimes (V \otimes V^*) \otimes (W \otimes W^*)$.

Claim. $M_{<l,m,n>} = Id_U \otimes Id_V \otimes Id_W = Id_{U\otimes V\otimes W}$

Proof. Consider $M_{<l,m,n>} : (U^* \otimes V)^* \times (V^* \otimes W)^* \to W^* \otimes U$ mapping $(v^* \otimes \bar{u}, w^* \otimes \bar{v})$ to $v^*(\bar{v})w^* \otimes \bar{u}$ for any $\bar{u} \in U, v^* \in V^*, \bar{v} \in V$ and $w^* \in W^*$. (Since in basis, it corresponds $a_{\bar{u}}b_{v^*}^T c_{\bar{v}}d_{w^*}^T = (b_{v^*}^T c_{\bar{v}})a_{\bar{u}}d_{w^*}^T$ here $a_{\bar{u}}$ is the corresponding column vector for $\bar{u}$ and $b_{v^*}^T$ is the row vector for v^*, and $b_{v^*}^T c_{\bar{v}}$ means $v^*(\bar{v})$.)

So for $M_{<l,m,n>} : (U^* \otimes V)^* \times (V^* \otimes W)^* \times (W^* \otimes U)^* \to \mathbb{C}$, it maps $(v^* \otimes \bar{u}, w^* \otimes \bar{v}, u^* \otimes \bar{w})$ to $u^*(\bar{u})v^*(\bar{v})w^*(\bar{w}) \in \mathbb{C}$, thus maps $v^* \otimes \bar{u} \otimes w^* \otimes \bar{v} \otimes u^* \otimes \bar{w}$ to $u^*(\bar{u})v^*(\bar{v})w^*(\bar{w})$. Thus $M_{<l,m,n>} = Id_U \otimes Id_V \otimes Id_W$ (In basis, send $u_i \otimes u_j^* \otimes v_k \otimes v_l^* \otimes w_s \otimes w_t^*$ to $\delta_{ij}\delta_{kl}\delta_{st}$, thus $M_{<l,m,n>} = \sum_{i,j}\sum_{k,l}\sum_{s,t}\delta_{ij}\delta_{kl}\delta_{st}u_i^* \otimes u_j \otimes v_k^* \otimes v_l \otimes w_s^* \otimes w_t = \sum_{i,k,s} u_i^* \otimes u_i \otimes v_k^* \otimes v_k \otimes w_s^* \otimes w_s$, and $Id_U = \sum_i u_i \otimes u_i^*$). □

Studying the complexity for matrix multiplication turns to study the border rank for the tensor $M_{<l,m,n>} = Id_U \otimes Id_V \otimes Id_W$. And by the below theorems, rank and border rank are both good measures for complexity.

Theorem. (Strassen[1])$\mathbf{R}(M_{<n>}) = O(n^w)$

Theorem. (Bini[3])$\underline{\mathbf{R}}(M_{<n>}) = O(n^w)$

$w := \inf\{h \in \mathbb{R} | n \times n$ matrices may be multiplied using $O(n^h)$ arithmetic operations. $\}$

For any efficient matrix multiplcation algorithm, the total complexity is determined by the number of scalar multiplications.[2], so rank and boder rank are both good measurement for matrix multiplication.

2. GEOMETRIC CONFIGURATIONS IN REDUCED MATRIX MULTIPLICATION ALGORITHMS

2.1 Border rank bounds for BCLRS tensors

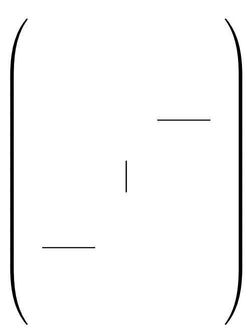
Define the generalized Bini-Capovani-Lotti-Romani-Smirnov tensor.

$$T_{BCLRS,m} := M_{<m,2,2>} - x_1^1 \otimes (y_1^2 \otimes z_1^2 + y_1^1 \otimes z_1^1)$$

Here $M_{<m,2,2>} = \sum_{i=1}^{m} \sum_{j=1,2} \sum_{k=1,2} x_j^i y_k^j z_j^k$, and the $T_{BCLRS,m}$ corresponds the $m \times 2$ and 2×2 matrix multiplication with entry x_1^1 equals 0. The observation below enables us to prove upper bounds for $\underline{\mathbf{R}}(M_{<m,2,2>})$ by finding the upper bounds for $\underline{\mathbf{R}}(T_{BCLRS,m})$.

Observation. [4] If $\underline{\mathbf{R}}(T_{BCLRS,m}) = r$ and $\underline{\mathbf{R}}(T_{BCLRS,m'}) = r'$, then for $n = m+m'-1$, $\underline{\mathbf{R}}(M_{<n,2,2>}) \leq r + r'$.

To see this, assume in $T_{BCLRS,m}$, we consider the matrix multiplication AB with A of size $m \times 2$ and B of size 2×2 . Here for A, entry x_2^m is missing. And in $T_{BCLRS,m'}$, assume the entry x_1^1 is missing. We put these two incomplete matrices together, we will get one with size $m + m' - 1$. Just like below,



(2.1)

Then we just combine these two algorithms together, we will get a algorithm for matrix multiplication tensor of $(m+m'-1)\times 2$ and 2×2.

By now, there exists algorithms for $T_{BCLRS,m}$ when $m = 2,3,4$, they show $\underline{\mathbf{R}}(T_{BCLRS,m}) \leq 3m-1$. If we could generalize the results when m is bigger, the claim above tells us that $\underline{\mathbf{R}}(M_{<m,2,2>}) \leq 3m+1$, which would be an improvement for the existing upper bounds $\underline{\mathbf{R}}(M_{<m,2,2>}) \leq 3m + \lfloor \frac{m}{7} \rfloor$. In the existing algorithms, there are good configurations hidden in these good algorithms. Next, we will analyze these algorithms and trying to find the similar algorithms when m is large.

2.2 Geometric notations for border rank algorithm

Definition. Let A,B,C be three vector spaces, $Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C) := \mathbb{P}\{T \in A\otimes B\otimes C|$ there exist $a\in A, b\in B, c\in C$ such that $T = a\otimes b\otimes c\} \subset \mathbb{P}(A\otimes B\otimes C)$. It is the set of rank one tensors up to scalar.

And there are only three types of lines on Segre variety $Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C)$: α-line is of the form $\mathbb{P}(< a_1, a_2 > \otimes b\otimes c)$ for some $a_1, a_2 \in A$ and $b\in B, c\in C$. And β-line and γ-line are defined similarly. For the matrix multiplication tensor $M_{<U,V,W>}$, A, B and C corresponds $U^*\otimes V, V^*\otimes W$ and $W^*\otimes U$. And there are two special types of α-lines, one is called the (α, u^*)-line and another one is (α, v)-line, where (α, u^*)-line is of the form $\mathbb{P}((< u_1^*, u_2^* >)\otimes b\otimes c >)$. Notice besides these two types, there are other $\alpha-$ lines, like $\mathbb{P}((su_1^*+tu_2^*)\otimes(sv_1+tv_2)\otimes b\otimes c)$.

A border rank algorithm for matrix multiplication is given in the form

$$T = \lim_{t\to 0} T_1(t) + \cdots + T_r(t)$$

with $T_i(t) \in$ is of rank one for $t\neq 0$. $\{T_i(t)\}$ are linear independent when $t\neq 0$. Use E_t to denote the r-plane spanned by $\{T_1(t), \ldots T_r(t)\}, E_t \in G(r, A\otimes B\otimes C)$. E denotes the limit r-plane of E_t when t goes to 0. T lies in E. The configurations arising in border rank algorithms for $T_{<BCLRS,m>}$ are interesting.

2.3 Configuration for $T_{BCLRS,2}$

The slightly modified reduced BCLR algorithm is given below. Notice that in the given algorithm, one point is stationary.

$$p_1(t) = x_2^1 \otimes (y_2^2 + y_1^2) \otimes (z_2^2 + tz_1^1)$$

$$p_2(t) = -(x_2^1 - tx_2^2) \otimes y_2^2 \otimes (z_2^2 + t(z_1^1 + z_1^2))$$

$$p_3(t) = x_1^2 \otimes (y_1^2 + ty_2^1) \otimes (z_2^2 + z_2^1)$$

$$p_4(t) = (x_1^2 - tx_2^2) \otimes (-y_1^2 + ty_1^1 - ty_2^1) \otimes z_2^1$$

$$p_5(t) = -(x_1^2 + x_2^1) \otimes y_1^2 \otimes z_2^2$$

$T_{BCLRS,2} = \lim_{t\to 0} \frac{1}{t}[p_1(t) + \cdots + p_5(t)]$

Consider the limit 5-plane $E := \lim_{t\to 0} < p_1(t), \ldots, p_5(t) > \in G(5, A \otimes B \otimes C)$, and $T_{BCLRS,2}$ is inside the 5-plane. Write $p_i = p_i(0)$, we have (up to scalar);

$$p_1 = x_2^1 \otimes (y_2^2 + y_1^2) \otimes z_2^2$$

$$p_2 = x_2^1 \otimes y_2^2 \otimes z_2^2$$

$$p_3 = x_1^2 \otimes y_1^2 \otimes (z_2^2 + z_2^1)$$

$$p_4 = x_1^2 \otimes y_1^2 \otimes z_2^1$$

$$p_5 = (x_1^2 + x_2^1) \otimes y_1^2 \otimes z_2^2$$

$S := \{p_1, p_2, p_3, p_4\}$ are linear independent and the limit 5-plane is spanned by $S \cup \{T_{BCLRS,2}\}$.

Claim. [5] The intersection of the limiting plane E and $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is the union of three lines. They are a special (β, w)-line $L_{12,(\beta,w)} = x_2^1 \otimes (v^2 \otimes W) \otimes z_2^1$, a special (γ, w^*)-line $L_{21,(\gamma,w^*)} = x_1^2 \otimes y_2^2 \otimes (W^* \otimes u_2)$ and a special α-line $L_\alpha = (x_2^1 + x_2^2) \otimes y_2^2 \otimes z_2^1$. The special line L_α is the unique line lying in $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ and intersecting with both $L_{12,(\beta,w)}$ and $L_{21,(\gamma,w^*)}$. Two points p_1, p_2 lie on the line $L_{12,(\beta,w)}$, the other two points p_3, p_4 are on $L_{21,(\gamma,w^*)}$, the stationary point p_5 lies on L_α.

Proof. Easy to check $p_1, p_2 \in L_{12,(\beta,w)}$ and $L_{12,(\beta,w)}$ is the unique line in the *Segre* connecting p_1 and p_2. Thus $L_{12,(\beta,w)} \subset E \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$. Similarly, $L_{21,(\gamma,w^*)}$ is the unique line on the Segre through points p_3 and p_4. Then any line in Segre connecting one point of $L_{21,(\gamma,w^*)}$ and another point on $L_{12,(\beta,w)}$ is in the span of set S, thus in the intersection. Easy to check the unique connecting line is just L_α. Thus these three lines are in the intersection, furthermore the three lines are exactly the intersection of span$S$ and the Segre.

Next need to show that no other points are in the intersection. Since E is the span of $S \cup \{T_{BCLRS,2}\}$, any point q in E is of the form $\lambda_1 q_1 + \lambda_2 T_{BCLRS,2}$ with q_1 ib the span of S. Notice that if $q \in Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$, then $\lambda_2 = 0$. If not, the coefficients of term $x_2^2 \otimes (y_2^2 \otimes z_2^2 + y_1^2 \otimes z_2^1)$ is $\lambda_2 \neq 0$ in q. Since rank of $x_2^2 \otimes (y_2^2 \otimes z_2^2 + y_1^2 \otimes z_2^1)$ is 2, thus q can not lie in the Segre. Thus $\lambda_2 = 0$, $E \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) =$ span of S $\cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$= the three lines. □

2.4 Configuration for $T_{BCLRS,3}$

Below is an algorithm for $T_{BCLRS,3}$ with one stationary point, it is a slightly modified version of *Alekseev* − *Smirnov* for $T_{BCLRS,3}$. Comparing the algorithm in

[6], here we make a linear transformation to make $p_8(t)$ a stable point.

$$p_1(t) = (-\frac{1}{2}t^2x_2^3 - \frac{1}{2}tx_2^2 + x_1^2) \otimes (-y_1^2 + y_2^2 + ty_1^1) \otimes z_3^1$$
$$p_2(t) = (t^2x_2^3 + tx_1^3 - \frac{1}{2}tx_2^2 - x_1^2) \otimes (y_1^2 + y_2^2 + ty_2^1) \otimes (z_3^2 - tz_2^2)$$
$$p_3(t) = (\frac{1}{2}t^2x_2^3 - tx_1^3 - \frac{1}{2}tx_2^2 + x_1^2) \otimes (y_1^2 + y_2^2 - ty_1^1) \otimes (z_3^1 - tz_2^1)$$
$$p_4(t) = (\frac{1}{2}tx_2^2 + x_1^2) \otimes (-y_1^2 + y_2^2 + ty_2^1) \otimes z_3^2$$

$$p_5(t) = (-t^2x_2^3 + tx_2^2 - x_2^1) \otimes y_1^2 \otimes (z_3^2 + \frac{1}{2}tz_2^1 - \frac{1}{2}tz_2^2 - t^2z_1^1)$$
$$p_6(t) = (tx_2^2 + x_2^1) \otimes y_2^2 \otimes (z_3^1 - \frac{1}{2}tz_2^1 + \frac{1}{2}tz_2^2 + t^2z_1^2)$$

$$p_7(t) = (-tx_1^3 + x_1^2 + \frac{1}{2}x_2^1) \otimes (y_1^2 + y_2^2) \otimes (-z_3^1 + tz_2^1 + z_3^2 - tz_2^2)$$
$$p_8(t) = (x_1^2 + \frac{1}{2}x_2^1) \otimes (y_1^2 - y_2^2) \otimes (z_3^1 + z_3^2)$$

And

$$T_{BCLRS,3} = \lim_{t\to 0} \frac{1}{t^2}[p_1(t) + \cdots + p_8(t)]$$

Claim. [6] Define $E := \lim_{t\to 0} < p_1(t), \ldots, p_8(t) > \in G(8, A \otimes B \otimes C)$. In the algorithm, $E \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is the union of three variety.

The first one is a sub-Segre variety $Seg_{21,(\beta,w),(\gamma,w^*)} := [x_1^2] \otimes \mathbb{P}(v^2 \otimes W) \otimes \mathbb{P}(W^* \otimes u_3)$, it is isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$.

The second one is a plane conic curve $C_{12,(\beta,w),(\gamma,w^*)} := \mathbb{P}(\cup_{[s,t]\in\mathbb{P}^1} x_2^1 \otimes (sy_1^2 - ty_2^2) \otimes (sz_3^2 + tz_3^1))$, it is isomorphic to $\mathbb{P}^1$.

The third variety is a family of lines $L_\alpha := \mathbb{P}(\cup_{[\sigma,\tau]\in\mathbb{P}^1} \cup_{[s,t]\in\mathbb{P}^1} ((\sigma x_2^1 + \tau x_1^2) \otimes (sy_1^2 - ty_2^2) \otimes (sz_3^2 + tz_3^1))$. These lines pass through the conic curve and the sub-Segre variety. L_α is isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$.

Let p_i denote the initial point $p_i(0)$, points $\{p_1, p_2, p_3, p_4\}$ are on the sub-Segre $Seg_{21,(\beta,w),(\gamma,w^*)}$, points $\{p_5, p_6\}$ lie on the conic curve $C_{12,(\beta,w),(\gamma,w^*)}$ and the left two points p_7, p_8 are on lines L_α.

Proof. $\{p_1, \ldots, p_7\}$ are linearly independent and p_8 is in the span of $S := \{p_1, \ldots, p_7\}$. The limit 8-plane E is the span of $S \cup \{T_{BCLRS,3}\}$. Consider the initial points (up to scalar)

$$
\begin{aligned}
p_1 &= x_1^2 \otimes (-y_1^2 + y_2^2) \otimes z_3^1 \\
p_2 &= x_1^2 \otimes (y_1^2 + y_2^2) \otimes z_3^2 \\
p_3 &= x_1^2 \otimes (y_1^2 + y_2^2) \otimes z_3^1 \\
p_4 &= x_1^2 \otimes (-y_1^2 + y_2^2) \otimes z_3^2
\end{aligned}
$$

Easy to check the span of $\{p_1, p_2, p_3, p_4\}$ intersected with $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is just the sub-Segre $Seg_{21,(\beta,w),(\gamma,w^*)}$.

$$
\begin{aligned}
p_5 &= x_2^1 \otimes y_1^2 \otimes z_3^2 \\
p_6 &= x_2^1 \otimes y_2^2 \otimes z_3^1 \\
p_7 = (x_1^2 + \frac{1}{2}x_2^1) &\otimes (y_1^2 + y_2^2) \otimes (-z_3^1 + z_3^2)
\end{aligned}
$$

Notice $p_7' := 2(p_7 - x^2 - 1 \otimes (y_1^2 + y_2^2) \otimes (-z_3^1 + z_3^2)) - p_5 + p_8 = x_2^1 \otimes y_1^2 \otimes z_3^1 - x_2^1 \otimes y_2^2 \otimes z_3^2$ is also in E, actually span $\{p_5, p_6, p_7'\}$ intersects with $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is just the plane conic curve $C_{12,(\beta,w),(\gamma,w^*)}$. To see this, any point q in the intersection must be of the form $x_2^1 \otimes (\alpha_1 y_1^2 + \alpha_2 y_2^2) \otimes (\beta_1 z_3^2 + \beta_2 z_3^1)$, and q is in the span of $\{p_5, p_6, p_7'\}$, thus $\alpha_1\beta_2 + \alpha_2\beta_1 = 0$. There for the intersection is $C_{12,(\beta,w),(\gamma,w^*)}$.

By the same argument as before, any $q \in E \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is a linear combination of points in S (if the coefficient of $T_{BCLRS,3}$ is not zero, the point q can't be of rank one). Thus, q lies on the lines in the Segre connecting points from $Seg_{21,(\beta,w),(\gamma,w^*)}$ and $C_{12,(\beta,w),(\gamma,w^*)}$, which is just L_α. □

2.5 Similarities between above two border rank algorithms

Firstly, for every matrix entry appeared in the algorithm, its order (eg, if tx_2^2 appears in the algorithm, the order for entry x_2^2 is 2) follows some pattern, for $m = 2$,

$$\begin{pmatrix} \star & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Here $\star$ means x_1^1 is removed in the reduced matric multiplication tensor.

For $m = 3$,

$$\begin{pmatrix} \star & 0 \\ 0 & 1 \\ 1 & 2 \end{pmatrix} \times \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 2 & 1 & 0 \\ 2 & 1 & 0 \end{pmatrix}$$

So can we find similar algorithms with this pattern when $m \geq 4$?

Next, in the configuration when $m = 3$, the three varieties $C_{12,(\beta,w),(\gamma,w^*)}$, $Seg_{21,(\beta,w),(\gamma,w^*)}$ and L_α respectively are analogous to the lines $L_{12,(\beta,w)}$, $L_{21,(\gamma,w^*)}$ and lines L_α. In $m = 3$ case, the last two varieties are isormorphic to $\mathbb{P}^1 \times \mathbb{P}^1$ and for the case $m = 2$, the first two lines are isomorphic to $\mathbb{P}^1$. In these two cases, L_α are lines in the Segre connecting points in these two varieties. So for large m, do there exist algorithms having similar ocnfigurations?

Notice that in $m = 2$ case, there is a stationary point lying on L_α. So can we improve the exsiting algorithm of $T_{BCLRS,3}$ such that only one stationary point lies on L_α, three limiting points are on the conic curve and the left four points are on the sub-Segre variety?

In the last part, we will give partial answers to these questions.

3. CONCLUSIONS

3.1 Nonexistence of certain improvement for algorithm in $T_{BCLRS,3}$

Assume the desired algorithm is of order 2. And the intersection of the limit 8-plane of the desired algorithm and the Segre variety $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ consists of two irreducible algebraic varieties, and lines connecting the points of these two surfaces. The first surface is

$$Seg_{21,(\beta,w),(\gamma,w^*)} := [x_1^2] \times \mathbb{P}(v^2 \otimes W) \times \mathbb{P}(W^* \otimes u_3)$$

The second curve is

$$C_{12,(\beta,w),(\gamma,w^*)} := \mathbb{P}(\cup_{[s,t]\in\mathbb{P}^1}\ x_2^1 \otimes (sy_1^2 - ty_2^2) \otimes (sz_3^2 + tz_3^1))$$

The family of lines $\mathbb{L}_\alpha$ is

$$\mathbb{L}_\alpha := \mathbb{P}(\cup_{[\sigma,\tau]\in\mathbb{P}^1} \cup_{[s,t]\in\mathbb{P}^1} (\sigma x_2^1 + \tau x_1^2) \otimes (sy_1^2 - ty_2^2) \otimes (sz_3^2 + tz_3^1))$$

In the desired algorithm, four initial points $p_1(0), p_2(0), p_3(0), p_4(0)$ are on $Seg_{21,(\beta,w),(\gamma,w^*)}$, three initial points $p_5(0), p_6(0), p_7(0)$ lie on the plane conic $C_{12,(\beta,w),(\gamma,w^*)}$ and the last stationary point p_8 lies on $\mathbb{L}_\alpha$ but not on C or on Seg_{21}.

If we want to make it a reduced border rank algorithm, we need point $p_i(t)$ is of rank 1 ($i = 1, \dots, 8$), and satisfy

$$T_{BCLR,3} = \lim_{t\to 0} \frac{1}{t^2}[p_1(t) + \cdots + p_8(t)]$$

where $T_{BCLR,3} = \sum_{i=1,2,3} \sum_{j=1,2} \sum_{k=1,2} x_j^i \otimes y_k^j \otimes z_i^k - x_1^1 \otimes y_1^1 \otimes z_1^1 - x_1^1 \otimes y_2^1 \otimes z_1^2$.

Since the algorithm is of order 2 and with only one stationary point, so the term $x_1^2 \otimes y_1^1 \otimes z_2^1 + x_1^2 \otimes y_2^1 \otimes z_2^2$ must come from the t^2 coefficient of the sum of four points on $Seg_{21,(\beta,w),(\gamma,w^*)}$. From the assumptions, only considering terms dealing with $x_1^2 \otimes y_1^1 \otimes z_2^1 + x_1^2 \otimes y_2^1 \otimes z_2^2$ we can set

$$\begin{pmatrix} p_1(0) \\ p_2(0) \\ p_3(0) \\ p_4(0) \end{pmatrix} = x_1^2 \otimes (\mathbf{a}y_1^2 + \mathbf{b}y_2^2 + \lambda^{\mathbf{1}} y_1^1 t + \lambda^{\mathbf{2}} y_2^1 t) \otimes (\mathbf{c}z_3^1 + \mathbf{d}z_3^2 + \mu^{\mathbf{1}} z_2^1 t + \mu^{\mathbf{2}} z_2^2 t))$$

and

$$p_8 = (\mathbf{s}x_2^1 + x_1^2) \otimes (\eta y_1^2 - \xi y_2^2) \otimes (\eta z_3^2 + \xi z_3^1)$$

with $a, b, c, d, \lambda^1, \lambda^2, \mu^1, \mu^2, \eta, \xi \in \mathbb{C}^4$ and ξ, η can not both be zero.

For the five points, the constant term and the term t should sum up to zero. And the t^2 term should sum up to $x_1^2 \otimes y_1^1 \otimes z_2^1 + x_1^2 \otimes y_2^1 \otimes z_2^2$, so we can set up equations.

For the t^2 term,

$$\lambda^1 \cdot \mu^1 = 1 \tag{3.1}$$

$$\lambda^2 \cdot \mu^2 = 1 \tag{3.2}$$

$$\lambda^1 \cdot \mu^2 = 0 \tag{3.3}$$

$$\lambda^2 \cdot \mu^1 = 0 \tag{3.4}$$

For the t term, with $k = 1, 2$

$$a \cdot \mu^k = 0 \tag{3.5}$$

$$b \cdot \mu^k = 0 \tag{3.6}$$

$$\lambda^k \cdot c = 0 \tag{3.7}$$

$$\lambda^k \cdot d = 0 \tag{3.8}$$

For the constant term,

$$a \cdot d = \eta^2 \tag{3.9}$$

$$a \cdot c = \eta\xi \tag{3.10}$$

$$b \cdot c = -\xi^2 \tag{3.11}$$

$$b \cdot d = -\eta\xi \tag{3.12}$$

where $a \cdot b = \sum_{i=1}^{4} a_i b_i$, and $a = (a_1, a_2, a_3, a_4), b = (b_1, b_2, b_3, b_4)$

Claim that either $\{a, b\}$ or $\{c, d\}$ is linear dependent.

If not, then both $\{a, b\}$ and $\{c, d\}$ are linear independent. From (1) (3), $\{\mu^1, \mu^2\}$ are linear independent. From (5) (6), span$\{a, b\}$ is perpendicular to span $\{\mu^1, \mu^2\}$, and by linear independence, we have $span\{a, b\}^\perp = span\{\mu^1, \mu^2\}$. Similarly, $span\{c, d\}^\perp = span\{\lambda^1, \lambda^2\}$.From (9)–(12), $\xi d - \eta c \in span\{a, b\}^\perp$. So $\xi d - \eta c = \alpha\mu^1 + \beta\mu^2$, by (1)–(4) $\alpha = (\xi d - \eta c) \cdot \lambda^1 = 0$ and $\beta = 0$ too. Thus $\xi d - \eta c = 0$ which contradicts that they are linearly independent.

Assume $\{c, d\}$ is linearly dependent, and since the group action of $GL(W)$ will still keep the algorithm and the configuration (since for any $A \in GL(W)$, $A \cdot (sw_1 - tw_2) \otimes (sw^2 + tw^1) = (sw_1 - tw_2) \otimes (sw^2 + tw^1)$), thus we can assume $d = 0$, now it is easy to see the four initial points $p_i(0)$ are linearly dependent, thus in this case the configuration would not be $Seg_{21,(\beta,w),(\gamma,w^*)}$. This means the desired algorithm would not exist.

3.2 Generalizing the algorithm to case $m = 4$

For the existing algorithm when $m = 4$, the geometry configuration is not good. So I am trying to come up with an algorithm with similar configurations when $m = 4$.

I start to assume the algorithm is of order 3 and is following the pattern below:

$$\begin{pmatrix} \star & 0 \\ 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{pmatrix} \times \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 3 & 2 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{pmatrix}$$

I am trying to find 11 points in the Segre variety such that

$$T_{BCLRS,4} = \lim_{t \to 0} \frac{1}{t^3}(T_1(t) + \cdots + T_{11}(t))$$

I am still in the progress trying to find the desired algorithm.

REFERENCES

[1] V. Strassen, *Gaussian elimination is not optimal*, *Numer. Math* **13**, 354-356 (1969)

[2] Peter Burgisser, Michael Clausen, and M. Amin Shokrollahi, *Algebraic complexity thoery*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematics Sciences], vol. 315, Springer- Verlag, Berlin, 1997, With the collection of Thomas Lickteig. MR. 99c:68002

[3] D.Bini, *Relations between exact and approximate bi-linear algorithms*, *Applications*, Calcolo **17** (1980), no. 1,87–97. MR 605920 (83f:68043b)

[4] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti, $O(n^{2.7799})$ *complexity for* $n \times n$ *approximate matrix multiplication*, Inform. Process. Lett. **8** (1979), no. **5**, 234-235. MR MR534068 (80h:68024)

[5] J.M Landsberg and N. Ryder, *On the geometry of border rank algorithms for* $n \times 2$ *by* 2×2 *matrix multiplication* **arXiv:**1509.08323v1 [cs.NA] 28 Sep 2015

[6] J.M Landsberg, 2011: *Tensors: Geometry and Applications* American Mathematics Society, Providence, Rhode Island, 439pp