Redesigning Strassen's Algorithm

Pádraig Ryan

UCD

July 2021

Outline

1. Notation
2. Tensors
    - Tensor Rank
3. Strassen’s Algorithm
4. Grochow-Moore
5. Search for Finite Field Vectors
6. Results
7. Further Questions

Notation

For the purposes of this presentation:

- Vectors, $v \in V$, in are written as column vectors.
- Dual Space covectors, $\nu \in V^*$ are written as row vectors.
- For the most part, we are working over finite fields, where vector spaces are isomorphic to their dual space.
- Finite fields are denoted by $\mathbb{F}$, and their n-dimensional vector space as $\mathbb{F}^n$.
- $(\mathbb{F}^n)^{\otimes m} = \underbrace{\mathbb{F}^n \otimes \mathbb{F}^n \otimes \ldots \otimes \mathbb{F}^n}_{m \text{ times}}$

Tensor Product

Definition (Tensor Product (of Vector Spaces))

Let V be a $\mathbb{F}$-Vector Space with basis $\{e_1, \ldots, e_n\}$ and W be a $\mathbb{F}$-Vector Space with basis $\{f_1, \ldots, f_m\}$. The tensor product of Vector Spaces V and W, denoted $V \otimes W$, is an *nm*-dimensional Vector Space with basis $\{e_i \otimes f_j : i \leq n, j \leq m\}$.

Example

Consider the Vector Spaces $\mathbb{F}^2$ and $(\mathbb{F}^3)^*$ with the standard basis vectors. $U := \mathbb{F}^2 \otimes (\mathbb{F}^3)^*$ is a $2 \cdot 3 = 6$ dimensional vector space with basis vectors:

$$e_1 \otimes f_1^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, e_1 \otimes f_2^* = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, e_1 \otimes f_3^* = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$e_2 \otimes f_1^* = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, e_2 \otimes f_2^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, e_2 \otimes f_3^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Tensors

A tensor describes a multilinear map between vector spaces.

Example (Matrix Multiplication Tensor)

The 3-slices of MM_2 are:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Tensor Rank

Definition

The **Rank** $rk(t)$ of a tensor $t \in (\mathbb{F}^n)^{\otimes n}$ is the minimum number, r, of tuples of n vectors $(u_i^1, \ldots, u_i^n)$ such that;

$$\sum_{i=1}^{r} u_i^1 \otimes \ldots \otimes u_i^n = t$$

Note: The bilinear complexity of a bilinear map is equivalent to the minimum Rank of a tensor.

Example

$$t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

has tensor rank $rk(t) = 2$, as;

$$t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

Matrix Multiplication and Strassen's Algorithm

Example (Matrix Multiplication Tensor)

The 3-slices of MM_2 are:

$$\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\begin{bmatrix}0&1&0&0\\0&0&0&1\\0&0&0&0\\0&0&0&0\end{bmatrix},\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&0&0&0\\0&0&1&0\end{bmatrix},\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$$

It can be checked that the rank of the Matrix multiplication tensor is no more than 8.

Volker Strassen Showed in 1968 that multiplying two 2×2 matrices could be done in 7 multiplications.

Strassen's Algorithm

Example (Strassen's Algorithm)

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$$\begin{aligned}
I &= (A_{11} + A_{22})(B_{11} + B_{22}) \\
II &= (A_{21} + A_{22})B_{11} \\
III &= A_{11}(B_{12} - B_{22}) \\
IV &= A_{22}(-B_{11} + B_{21}) \\
V &= (A_{11} + A_{12})B_{22} \\
VI &= (-A_{11} + A_{21})(B_{11} + B_{12}) \\
VII &= (A_{12} - A_{22})(B_{21} + B_{22})
\end{aligned}$$

Example (Strassen's Algorithm)

$$I = (A_{11} + A_{22})(B_{11} + B_{22})$$

$$II = (A_{21} + A_{22})B_{11}$$

$$III = A_{11}(B_{12} - B_{22})$$

$$IV = A_{22}(-B_{11} + B_{21})$$

$$V = (A_{11} + A_{12})B_{22}$$

$$VI = (-A_{11} + A_{21})(B_{11} + B_{12})$$

$$VII = (A_{12} - A_{22})(B_{21} + B_{22})$$

$$AB = \begin{bmatrix} I + IV - V + VII & III + V \\ II + IV & I + III - II + VI \end{bmatrix}$$

Strassen's Algorithm as a Tensor

We can see that Strassen's Algorithm gives the MM_2 tensor as follows; Consider C_{11}.

$$C_{11} = I + IV - V + VII$$

$$(M_{ij1})_{i,j} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \otimes \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} \otimes \begin{bmatrix}-1 & 0\\1 & 0\end{bmatrix} + (-1)\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix} \otimes \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$$

$$+ \begin{bmatrix}0 & 1\\0 & -1\end{bmatrix} \otimes \begin{bmatrix}0 & 0\\1 & 1\end{bmatrix}$$

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

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

Grochow-Moore

Designing Strassen's Algorithm (Grochow-Moore) describes a way in which Strassen's Algorithm can be generalised to vectors of length n.

Definition (Unitary 2-design)

A set, S, of n-dimensional vectors is a unitary 2-design if;

$$\sum_{v \in S} v = 0 \text{ and } \frac{1}{|S|} \sum_{v \in S} v \otimes v^* = \frac{1}{n}\mathbf{1}$$

Theorem

Let $S = \{w_1, \ldots, w_s\} \subset \mathbb{C}^n$ be a unitary 2-design, and let $s = |S|$. Then the tensor rank of MM_n is at most $s(s-1)(s-2)+1$, and the rank 1 decomposition is given by:

$$MM_n = \mathbf{1}^{\otimes 3} + \frac{n^3}{s^3} \sum_{i,j,k,distinct} (w_i \otimes (w_j - w_i)^*) \otimes (w_j \otimes (w_k - w_j)^*) \otimes (w_k \otimes (w_i - w_k)^*)$$

Example

The vertices of an equilateral triangle (centered at the origin) form a 2-design.

$$S := \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}, \begin{bmatrix} -1/2 \\ -\sqrt{3}/2 \end{bmatrix} \right\}$$

$$\sum_{v \in S} v = \begin{bmatrix} 1 - 1/2 - 1/2 \\ \sqrt{3}/2 - \sqrt{3}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

$$\frac{1}{|S|} \sum_{v \in S} v \otimes v^*$$

$$= \frac{1}{3} \cdot \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{bmatrix} + \begin{bmatrix} 1/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 3/4 \end{bmatrix} \right) = \frac{1}{2}\mathbf{1}$$

Hence, the rank of $MM_2 = s(s-1)(s-2) + 1 = 7$

Search for Finite Field Vectors

The above construction for MM_2 will only work in fields where $\frac{1}{2}$ and $\sqrt{3}$ exist.
Can we find vectors that satisfy some version of Grochow-Moore's construction over finite fields or over the Integers?
Hence lets consider:

$$\begin{aligned} MM_n - \mathbf{1}^{\otimes 3} = \lambda \sum_{g \in G} [g_1(w_1) \otimes (g_2(w_2))^*] \otimes [g_3(w_3) \otimes (g_4(w_4))^*] \\ \otimes [g_5(w_5) \otimes (g_6(w_6))^*] \\ \in (\mathbb{F}^{n \times n})^{\otimes 3} \end{aligned}$$

with subgroup $G \leq GL(n, \mathbb{F})^6$.
Note: the tensor rank is invariant under the action of elements of the General Linear group.

We would like to work in $(\mathbb{F}^n)^{\otimes 6}$ rather than $(\mathbb{F}^{n\times n})^{\otimes 3}$.

Definition

We define an isomorphism $\pi : (\mathbb{F}^{n\times n})^{\otimes 3} \mapsto (\mathbb{F}^n)^{\otimes 6}$

$$\pi\left(e_{i_1} \otimes e^*_{i_2} \otimes e_{i_3} \otimes e^*_{i_4} \otimes e_{i_5} \otimes e^*_{i_6}\right) = e_{i_1} \otimes e_{i_2} \otimes e_{i_3} \otimes e_{i_4} \otimes e_{i_5} \otimes e_{i_6}$$

Let $g = (g_1, g_2, g_3, g_4, g_5, g_6) \in GL(n, \mathbb{F})^6$

$$\pi\left(g_1(e_{i_1}) \otimes (g_2(e_{i_2}))^* \otimes g_3(e_{i_3}) \otimes (g_4(e_{i_4}))^* \otimes g_5(e_{i_5}) \otimes (g_6(e_{i_6}))^*\right) = g_1(e_{i_1}) \otimes g_2(e_{i_2}) \otimes g_3(e_{i_3}) \otimes g_4(e_{i_4}) \otimes g_5(e_{i_5}) \otimes g_6(e_{i_6})$$

We want to pick a subgroup $G \leq GL(n, \mathbb{F})^6$ that fixes MM_n and $\mathbf{1}^{\otimes 3}$
Choosing g to fix $\mathbf{1}^{\otimes 3}$ and MM_n gets us $g = \left(g_1, g_1^{-1}, g_1, g_1^{-1}, g_1, g_1^{-1}\right)$

$$\pi([MM_n - \mathbf{1}^{\otimes 3}) = \pi\Bigg[\lambda \sum_{g \in G} [g_1(w_1) \otimes (g_2(w_2))^*] \otimes [g_3(w_3) \otimes (g_4(w_4))^*] \otimes [g_5(w_5) \otimes (g_6(w_6))^*]\Bigg]$$

$$= \lambda \sum_{g \in G} g_1(w_1) \otimes g_1^{-T}(w_2) \otimes g_1(w_3) \otimes g_1^{-T}(w_4) \otimes g_1(w_5) \otimes g_1^{-T}(w_6)$$

$$\in (\mathbb{F}^n)^{\otimes 6}$$

I choose the subgroup of $GL(n, \mathbb{F})$ to be S_{n+1}, the symmetric group on a set of $n+1$ elements.

Example (Representation of S_3 in $GL(2, \mathbb{F})$)

The subgroup S_3 is generated by

$$\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$$

Results of Magma Calculation

Note for $\mathbb{Z}$, we checked only vectors of the form $v \in \{-1, 0, 1\}^n$

n	$\mathbb{Z}$	$\mathbb{F}_2^n$	$\mathbb{F}_3^n$	$\mathbb{F}_4^n$	$\mathbb{F}_5^n$	$\mathbb{F}_7^n$	$\mathbb{F}_8^n$
2	6	6	6	6	12	12	6
3	24	24	48				

Further Questions

- Some finite fields have multiple sets of vectors which generate MM_n?
- The Symmetric group S_{n+1} has order $(n+1)!$. Hence it only gives the minimum tensor rank when $n = 2$. For $n = 3$, it gives an upper bound of 25 for the tensor rank. This is better than the naive approach, 27, but falls short of the current best upper bound of 23. Can picking different subgroups and group orbits get a better bound?
- Is there any way to generate the vectors without searching the tensor product space?