FLIP GRAPHS FOR MATRIX MULTIPLICATION

Jakob Moosbauer

Joint work with Manuel Kauers

WACT 2023

JOHANNES KEPLER
UNIVERSITY LINZ

Matrix Multiplication

- Matrix multiplication is an interesting problem.
- The standard algorithm for multiplying two $n \times n$ matrices uses $\mathrm{O}(n^3)$ operations.
- Strasen's algorithm uses a multiplication scheme for 2×2 matrices that needs only 7 multiplications instead of 8, yielding a complexity of $\mathrm{O}(n^{2.81})$
- Proving upper (or lower) bounds on the rank of specific tensors is hard.
- The exact rank is only known for multiplication of 2×2 by $2 \times n$ matrices.
- Strassen's algorithm is the only fast algorithm that can be used in practice.
- An algorithm found by AlphaTensor can multiply 4×4 matricers using only 47 multiplications. Sadly it only works over rings of characteristic 2.

Matrix Multiplication Schemes

$$
\begin{aligned}
m_1 &= a_{1,1}b_{1,1} \\
m_2 &= a_{1,2}b_{2,1} \\
m_3 &= a_{1,1}b_{1,2} \\
m_4 &= a_{1,2}b_{2,2} \\
m_5 &= a_{2,1}b_{1,1} \\
m_6 &= a_{2,2}b_{2,1} \\
m_7 &= a_{2,1}b_{1,2} \\
m_8 &= a_{2,2}b_{2,2} \\
c_{1,1} &= m_1 + m_2 \\
c_{1,2} &= m_3 + m_4 \\
c_{2,1} &= m_5 + m_6 \\
c_{2,2} &= m_7 + m_8
\end{aligned}
$$

Matrix Multiplication Schemes

$$
\begin{aligned}
m_1 &= a_{1,1}b_{1,1} \\
m_2 &= a_{1,2}b_{2,1} \\
m_3 &= a_{1,1}b_{1,2} \\
m_4 &= a_{1,2}b_{2,2} \\
m_5 &= a_{2,1}b_{1,1} \\
m_6 &= a_{2,2}b_{2,1} \\
m_7 &= a_{2,1}b_{1,2} \\
m_8 &= a_{2,2}b_{2,2} \\
c_{1,1} &= m_1 + m_2 \\
c_{1,2} &= m_3 + m_4 \\
c_{2,1} &= m_5 + m_6 \\
c_{2,2} &= m_7 + m_8
\end{aligned}
$$

$$
\begin{aligned}
m_1 &= (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2}) \\
m_2 &= (a_{1,1} + a_{1,2})(b_{2,2}) \\
m_3 &= (a_{2,1} + a_{2,2})(b_{1,1}) \\
m_4 &= (a_{1,1})(b_{1,2} - b_{2,2}) \\
m_5 &= (a_{2,2})(b_{2,1} - b_{1,1}) \\
m_6 &= (a_{2,1} - a_{1,1})(b_{1,1} + b_{1,2}) \\
m_7 &= (a_{1,2} - a_{2,2})(b_{2,1} + b_{2,2}) \\
c_{1,1} &= m_1 - m_2 + m_5 + m_7 \\
c_{1,2} &= m_2 + m_4 \\
c_{2,1} &= m_3 + m_5 \\
c_{2,2} &= m_1 - m_3 + m_4 + m_6
\end{aligned}
$$

Matrix Multiplication Schemes

$$
\begin{aligned}
m_1 &= (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2}) \\
m_2 &= (a_{1,1} + a_{1,2})(b_{2,2}) \\
m_3 &= (a_{2,1} + a_{2,2})(b_{1,1}) \\
m_4 &= (a_{1,1})(b_{1,2} - b_{2,2}) \\
m_5 &= (a_{2,2})(b_{2,1} - b_{1,1}) \\
m_6 &= (a_{2,1} - a_{1,1})(b_{1,1} + b_{1,2}) \\
m_7 &= (a_{1,2} - a_{2,2})(b_{2,1} + b_{2,2}) \\
c_{1,1} &= m_1 - m_2 + m_5 + m_7 \\
c_{1,2} &= m_2 + m_4 \\
c_{2,1} &= m_3 + m_5 \\
c_{2,2} &= m_1 - m_3 + m_4 + m_6
\end{aligned}
$$

Matrix Multiplication Schemes

$$
\begin{aligned}
m_1 &= (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2}) \\
m_2 &= (a_{1,1} + a_{1,2})(b_{2,2}) \\
m_3 &= (a_{2,1} + a_{2,2})(b_{1,1}) \\
m_4 &= (a_{1,1})(b_{1,2} - b_{2,2}) \\
m_5 &= (a_{2,2})(b_{2,1} - b_{1,1}) \\
m_6 &= (a_{2,1} - a_{1,1})(b_{1,1} + b_{1,2}) \\
m_7 &= (a_{1,2} - a_{2,2})(b_{2,1} + b_{2,2}) \\
c_{1,1} &= m_1 - m_2 + m_5 + m_7 \\
c_{1,2} &= m_2 + m_4 \\
c_{2,1} &= m_3 + m_5 \\
c_{2,2} &= m_1 - m_3 + m_4 + m_6
\end{aligned}
$$

$$
\begin{aligned}
(a_{1,1} + a_{2,2}) \otimes (b_{1,1} + b_{2,2}) \otimes (c_{1,1} + c_{2,2}) +& \\
(a_{1,1} + a_{1,2}) \otimes b_{2,2} \otimes (c_{1,2} - c_{1,1}) +& \\
(a_{2,1} + a_{2,2}) \otimes b_{1,1} \otimes (c_{2,1} - c_{2,2}) +& \\
a_{1,1} \otimes (b_{1,2} - b_{2,2}) \otimes (c_{1,2} + c_{2,2}) +& \\
a_{2,2} \otimes (b_{2,1} - b_{1,1}) \otimes (c_{1,1} + c_{2,1}) +& \\
(a_{2,1} - a_{1,1}) \otimes (b_{1,1} + b_{1,2}) \otimes c_{2,2} +& \\
(a_{1,2} + a_{2,2}) \otimes (b_{2,1} + b_{2,2}) \otimes c_{1,1}&
\end{aligned}
$$

Matrix Multiplication Schemes

$$
\begin{aligned}
(a_{1,1} + a_{2,2}) \otimes (b_{1,1} + b_{2,2}) \otimes (c_{1,1} + c_{2,2}) +\\
(a_{1,1} + a_{1,2}) \otimes b_{2,2} \otimes (c_{1,2} - c_{1,1}) +\\
(a_{2,1} + a_{2,2}) \otimes b_{1,1} \otimes (c_{2,1} - c_{2,2}) +\\
a_{1,1} \otimes (b_{1,2} - b_{2,2}) \otimes (c_{1,2} + c_{2,2}) +\\
a_{2,2} \otimes (b_{2,1} - b_{1,1}) \otimes (c_{1,1} + c_{2,1}) +\\
(a_{2,1} - a_{1,1}) \otimes (b_{1,1} + b_{1,2}) \otimes c_{2,2} +\\
(a_{1,2} + a_{2,2}) \otimes (b_{2,1} + b_{2,2}) \otimes c_{1,1}
\end{aligned}
$$

$$
\begin{aligned}
a_{1,1} \otimes b_{1,1} \otimes c_{1,1} +\\
a_{1,2} \otimes b_{2,1} \otimes c_{1,1} +\\
a_{1,1} \otimes b_{1,2} \otimes c_{1,2} +\\
a_{1,2} \otimes b_{2,2} \otimes c_{1,2} +\\
a_{2,1} \otimes b_{1,1} \otimes c_{2,1} +\\
a_{2,2} \otimes b_{2,1} \otimes c_{2,1} +\\
a_{2,1} \otimes b_{1,2} \otimes c_{2,2} +\\
a_{2,2} \otimes b_{2,2} \otimes c_{2,2}
\end{aligned}
$$

Matrix Multiplication Schemes

Definition

Let $n, m, p \in \mathbb{N}$. The matrix multiplication tensor is defined by

$$\mathcal{M}_{n,m,p} = \sum_{i,j,k=1}^{n,m,p} a_{i,j} \otimes b_{j,k} \otimes c_{k,i} \in K^{n,m} \otimes K^{m,p} \otimes K^{p,n}$$

where $a_{x,y}, b_{x,y}$ and $c_{x,y}$ refer to the matrices of the respective format that have a 1 at position (x, y) and zeros elsewhere.

Rank-one tensors are non-zero tensors of the form $A \otimes B \otimes C$.

The rank of a tensor $\mathcal{T}$ is the smallest number r such that $\mathcal{T}$ can be written as a sum of r rank one tensors.

An (n, m, p)-matrix multiplication scheme is a finite set S of rank one tensors, such that $\mathcal{M}_{n,m,p} = \sum_{t \in S} t$. We call $|S|$ the rank of the scheme.

Reductions

Under certain conditions some rank-one tensors can be combined leading to a reduction in the number of multiplications.

Consider the rank one tensors

$$a_{3,1} \otimes b_{1,2} \otimes c_{1,1}$$
$$a_{3,1} \otimes b_{1,2} \otimes c_{2,1}.$$

Their sum is again a rank one tensor:

$$a_{3,1} \otimes b_{1,2} \otimes (c_{1,1} + c_{2,1}).$$

For two rank one tensors such a combination is possible if and only if two of the factors are constant multiples of each other.

Reductions

It is sufficient that the second factors are linearly dependent, for example:

$$a_{1,1} \otimes b_{1,1} \otimes c_{1,1}$$
$$a_{1,1} \otimes b_{1,2} \otimes c_{3,1}$$
$$a_{1,1} \otimes (b_{1,1} + b_{1,2}) \otimes c_{2,2}.$$

These can be combined to

$$a_{1,1} \otimes b_{1,1} \otimes (c_{1,1} + c_{2,2})$$
$$a_{1,1} \otimes b_{1,2} \otimes (c_{3,1} + c_{2,2}).$$

Reductions

Definition

Let $n, m, p, r \in \mathbb{N}$ and let $S = \{A^{(i)} \otimes B^{(i)} \otimes C^{(i)} \mid i \in \{1, \ldots, r\}\}$ be an (n, m, p)-matrix multiplication scheme. We call S reducible if there is a nonempty set $I \subseteq \{1, \ldots, r\}$ such that

1. $\dim_K \langle A^{(i)} \rangle_{i \in I} = 1$ and
2. $\dim_K \langle B^{(i)} \rangle_{i \in I} < |I|$,

or analogously with B, A or A, C or C, A or B, C or C, B in place of A, B.

Proposition

Let $n, m, p, r \in \mathbb{N}$ and let S be a reducible (n, m, p)-matrix multiplication scheme of rank r. Then there exists an (n, m, p)-matrix multiplication scheme of rank $r-1$.

Symmetries

The group $G = GL_n(K) \times GL_m(K) \times GL_p(K)$ acts on a rank-one tensor $A \otimes B \otimes C \in K^{n,m} \otimes K^{m,p} \otimes K^{p,n}$ by

$$(U, V, W)(A \otimes B \otimes C) = UAV^{-1} \otimes VBW^{-1} \otimes WCU^{-1}$$

The matrix multiplication tensor is invariant under this action.

We call two matrix mutliplication schemes S_1 and S_2 equivalent if they belong to the same orbit.

- Since G acts linearly on A, B and C, reducibility is preserved by this action.
- We associate a matrix multiplication scheme with its equivalence class.
- For small matrices we can compute a normal form.

Flips

$A_1 \otimes B_1 \otimes C_1$

$A_1 \otimes B_2 \otimes C_2$

Flips

$$+A_1 \otimes B_1 \otimes C_2$$

$$A_1 \otimes B_1 \otimes C_1$$

$$A_1 \otimes B_2 \otimes C_2 \qquad \longrightarrow$$

$$-A_1 \otimes B_1 \otimes C_2$$

Flips

$$+A_1 \otimes B_1 \otimes C_2$$

$$\begin{matrix} A_1 \otimes B_1 \otimes C_1 \\ A_1 \otimes B_2 \otimes C_2 \end{matrix} \quad \longrightarrow \quad \begin{matrix} A_1 \otimes B_1 \otimes (C_1 + C_2) \\ A_1 \otimes (B_2 - B_1) \otimes C_2 \end{matrix}$$

$$-A_1 \otimes B_1 \otimes C_2$$

Flips

$$A_1 \otimes B_1 \otimes C_1 + A_2 \otimes B_2 \otimes C_1 =$$

$$A_1 \otimes (B_1 + B_2) \otimes C_1 + (A_2 - A_1) \otimes B_2 \otimes C_1 =$$

$$A_1 \otimes (B_1 - B_2) \otimes C_1 + (A_2 + A_1) \otimes B_2 \otimes C_1 =$$

$$(A_1 + A_2) \otimes B_1 \otimes C_1 + A_2 \otimes (B_2 - B_1) \otimes C_1 =$$

$$(A_1 - A_2) \otimes B_1 \otimes C_1 + A_2 \otimes (B_2 + B_1) \otimes C_1$$

Flips

Definition

Let $n, m, k, r \in \mathbb{N}$ and let S, S' be (n, m, p)-matrix multiplication schemes of rank r. We call S' a flip of S if there are

- $T_1 = A_1 \otimes B_1 \otimes C_1 \in S$,
- $T_2 = A_2 \otimes B_2 \otimes C_1 \in S$ and
- $T \in \{A_1 \otimes B_2 \otimes C_1, A_2 \otimes B_1 \otimes C_1\}$

such that $(S \setminus \{T_1, T_2\}) \cup \{T_1 + T, T_2 - T\} = S'$.
We also call S' a flip of S if the defintion applies analogously for a permutation of A, B and C.

The Flip Graph

Definition

Let $n, m, p \in \mathbb{N}$ and let V be the set of all orbits of (n, m, p)-matrix multiplication schemes under the symmetry group and define

$$E_1 = \{(S, S') \mid S' \text{ is a flip of } S\}$$
$$E_2 = \{(S, S') \mid S' \text{ is a reduction of } S\}.$$

1. The graph $G = (V, E_1 \cup E_2)$ is called the (n, m, p)-flip graph. The edges in E_1 are called flips and the edges in E_2 are called reductions.
2. For a given $r \in \mathbb{N}$, the set $\{S \in V : \mathrm{rank}(S) = r\}$ is called the rth level of G.

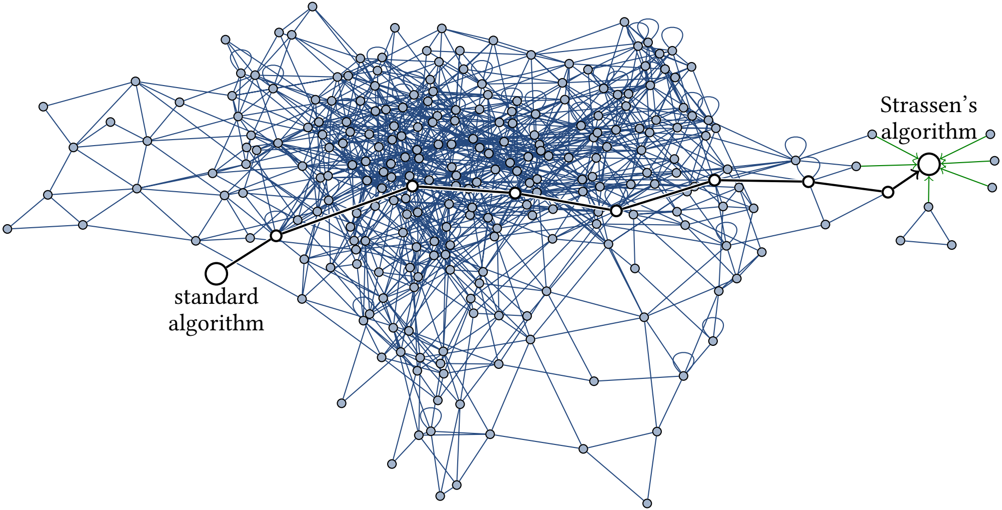
Flips

If we take $\mathbb{Z}_2$ as base field then there are only 2 possible flips for every pair of lines.

$$A_1 \otimes B_1 \otimes C_1 + A_2 \otimes B_2 \otimes C_1 =$$
$$A_1 \otimes (B_1 + B_2) \otimes C_1 + (A_2 + A_1) \otimes B_2 \otimes C_1 =$$
$$(A_1 + A_2) \otimes B_1 \otimes C_1 + A_2 \otimes (B_2 + B_1) \otimes C_1$$

The advantage is, that we get matching factors more often and the set of coefficients doesn't grow when we do a flip.

The Flip Graph



The Flip Graph

- 273 vertices
- 1183 edges
- 2 components
- length of the shortest path from the standard algorithm to Strassen: 8
- diameter: 12
- The same procedure is not duable for $3, 3, 3$-matrix multiplication
  - At distance 1 from the standard algorithm there is 1 vertex.
  - At distance 2 from the standard algorithm there are about 600 vertices.
  - At distance 3 from the standard algorithm there are about 20 000 vertices.
  - At distance 4 from the standard algorithm there are nearly 600 000 vertices.

Random Search

To find matrix multiplication schemes of lower rank we use the following search strategy:

Algorithm 1

Input: A matrix multiplication scheme S and a limit ℓ for the path length.

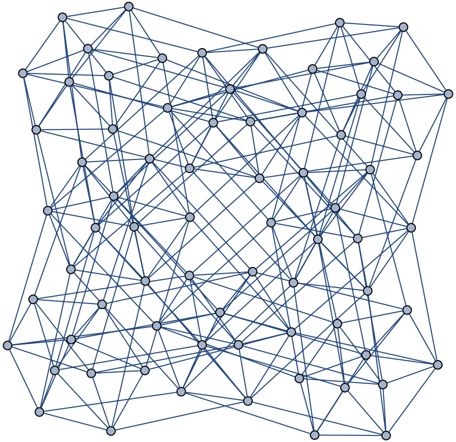
Output: A matrix multiplication scheme with rank decreased by one or $\perp$.

```
if S has no neighbours, return ⊥
for i = 1, ..., ℓ, do:
    if S is reducible, then return a reduction of S.
    if one of the neighbours of S is reducible, then return a reduction of it.
    Set S to a randomly selected neighbour of S.
return ⊥
```

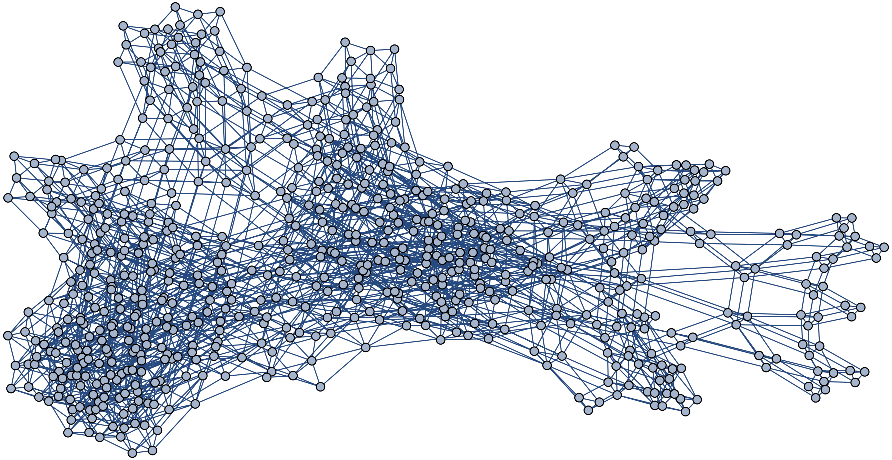
3×3 Matrices

- We can do a completion of the graph at the 23 multiplication level.
- In total we so find over 64 000 non-equivalent multiplication schemes.
- We identify 584 connected components.
- The smallest components are 40 isolated vertices.
- The largest component contains 6630 vertices.
- On average every vertex has 17 neighbours.

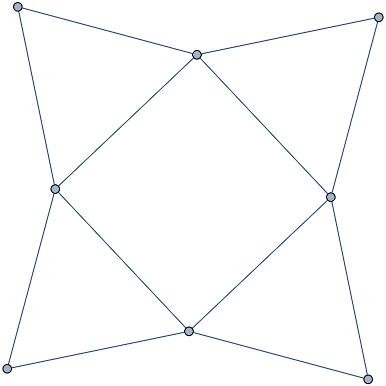
3×3 Matrices



3×3 Matrices



3×3 Matrices



Search Strategy

Algorithm 2

Input: A set P of schemes of a certain rank, a path length limit ℓ, a pool size limit s, and a target rank r

Output: A set Q of s schemes of rank r

```
if P consists of schemes of rank r, return P.
Q = ∅
while |Q| < s do:
    apply Alg. 1 to a random element of P and ℓ.
    if Alg. 1 returns a scheme, add it to Q.
call the algorithm recursively with Q in place of P.
```

Summary of Results

Size	Best known algorithm	Our algorithm
(2,2,2)	7	7
(2,2,3)	11	11
(2,2,4)	14	14
(2,2,5)	18	18
(2,3,3)	15	15
(2,3,4)	20	20
(2,3,5)	25	25
(2,4,4)	26	26
(2,4,5)	33	33
(2,5,5)	40	40
(3,3,3)	23	23
(3,3,4)	29	29
(3,3,5)	36	36
(3,4,4)	38	38
(3,4,5)	47	47
(3,5,5)	58	58
(4,4,4) mod 2	47	47
(4,4,4)	49	49
(4,4,5) mod 2	63	60
(4,4,5)	63	62
(4,5,5)	76	76
(5,5,5) mod 2	96	95
(5,5,5)	98	97

Size	Best known algorithm	Our algorithm
(2,2,6)	21	21
(2,3,6)	30	30
(2,4,6)	39	39
(2,5,6)	48	48
(2,6,6)	57	56
(3,3,6)	40	42
(3,4,6)	56	57
(3,5,6)	70	71
(3,6,6)	80	93
(4,4,6)	75	74
(4,5,6)	93	93
(5,5,6)	116	116
(6,6,6)	160	164

The Brent Equations

$$
\begin{aligned}
m_1 &= (\alpha^{(1)}_{1,1} a_{1,1} + \alpha^{(1)}_{1,2} a_{1,2} + \ldots)(\beta^{(1)}_{1,1} b_{1,1} + \beta^{(1)}_{1,2} b_{1,2} + \ldots) \\
&\vdots \\
m_r &= (\alpha^{(r)}_{1,1} a_{1,1} + \alpha^{(r)}_{1,2} a_{1,2} + \ldots)(\beta^{(r)}_{1,1} b_{1,1} + \beta^{(r)}_{1,2} b_{1,2} + \ldots) \\
c_{1,1} &= (\gamma^{(1)}_{1,1} m_1 + \ldots + \gamma^{(r)}_{1,1} m_r) \\
&\vdots \\
c_{n,p} &= (\gamma^{(1)}_{n,p} m_1 + \ldots + \gamma^{(r)}_{n,p} m_r)
\end{aligned}
$$

The coefficients need to be such that

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

The Brent Equations

$$\sum_{l=1}^{r} \alpha_{i_1,i_2}^{(l)} \beta_{j_1,j_2}^{(l)} \gamma_{k_1,k_2}^{(l)} = \delta_{i_2,j_1} \delta_{i_1,k_1} \delta_{j_2,k_2}$$

$$i_1, k_1 \in \{1, \ldots, n\}$$
$$i_2, j_1 \in \{1, \ldots, m\}$$
$$j_2, k_2 \in \{1, \ldots, p\}$$

System with $r(nm + mp + pn)$ variables and $n^2m^2p^2$ cubic equations.

Lifting solutions

To lift solutions modulo 2 to solutions over the integers we apply Hensel lifting. This allows us to lift a solution modulo 2 to a solution modulo 2^k.

Assume we have a solution modulo 2^k:

$$\sum_{l=1}^{r} \alpha_{i_1,i_2}^{(l)} \beta_{j_1,j_2}^{(l)} \gamma_{k_1,k_2}^{(l)} = \delta$$

We make the following ansatz modulo 2^{k+1}:

$$\sum_{l=1}^{r} (\alpha_{i_1,i_2}^{(l)} + 2^k \hat{\alpha}_{i_1,i_2}^{(l)})(\beta_{j_1,j_2}^{(l)} + 2^k \hat{\beta}_{j_1,j_2}^{(l)})(\gamma_{k_1,k_2}^{(l)} + 2^k \hat{\gamma}_{k_1,k_2}^{(l)}) = \delta$$

Completeness results

Theorem

The flip graph is weakly connected.

Theorem

Let $n, m, p, r \in \mathbb{N}$ and let S_1, S_2 be two irreducible (n, m, p)-matrix multiplication schemes of rank r over $K = \mathbb{Z}_2$. If S_1 and S_2 differ in exactly two elements, then S_1 is a flip of S_2.

Future work and open questions

- Improve search strategy (symmetries, structure, heuristics, machine learning)
- Identify good starting points
- Flips that modify more than 2 rank-one tensors
- Recognize/avoid local minima
- Other tensors to decompose
- Construct a path between to vertices
- Properties of the flip graph
- Larger ground fields
- Border rank
- Quadratic algorithms