Searching for Cyclic-Invariant Fast Matrix Multiplication Algorithms

João Pinheiro, Grey Ballard, Frank Moore

Department of Computer Science

WAKE FOREST UNIVERSITY

Searching for Fast Matrix Multiplication Algorithms

Classical Matrix Multiplication:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$AB = C, \mathcal{O}(n^3), \text{ for } A, B \in \mathbb{R}^{n\times n}$$

Fast algorithms pre-compute sums and differences of inputs and then use the distributive property of multiplication followed by more sums to carefully cancel terms [1]. Strassen's 1969 algorithm is perhaps the best example, which has a computational cost of $\mathcal{O}(n^{log_2 7}) \approx \mathcal{O}(n^{2.81})$.

Constructing a Matrix Multiplication Tensor

Matrix Multiplication as a Tensor Operation:

$$\mathcal{M} \times_1 vec(A) \times_2 vec(B) = \mathcal{M} \times_1 \begin{bmatrix} A_{11} \\ A_{12} \\ A_{21} \\ A_{22} \end{bmatrix} \times_2 \begin{bmatrix} B_{11} \\ B_{12} \\ B_{21} \\ B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} \\ C_{21} \\ C_{12} \\ C_{22} \end{bmatrix} = vec(C^{\mathsf{T}})$$

$$\mathrm{M}_1 = \begin{bmatrix} 1 & & & \\ & & 1 & \\ & & & \\ & & & \end{bmatrix} \quad \mathrm{M}_2 = \begin{bmatrix} & & & \\ & & & \\ 1 & & & \\ & & 1 & \end{bmatrix} \quad \mathrm{M}_3 = \begin{bmatrix} & 1 & & \\ & & & 1 \\ & & & \\ & & & \end{bmatrix} \quad \mathrm{M}_4 = \begin{bmatrix} & & & \\ & & & \\ & 1 & & \\ & & & 1 \end{bmatrix}$$

For Example:

$$\mathrm{M}_3 \times_1 \begin{bmatrix} A_{11} \\ A_{12} \\ A_{21} \\ A_{22} \end{bmatrix} \times_2 \begin{bmatrix} B_{11} \\ B_{12} \\ B_{21} \\ B_{22} \end{bmatrix} = [A_{11}\ A_{12}\ A_{21}\ A_{22}] \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{12} \\ B_{21} \\ B_{22} \end{bmatrix} = A_{11}B_{12} + A_{12}B_{22} = C_{12}$$

Strassen's Algorithm as Factor Matrices of a KTensor:

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

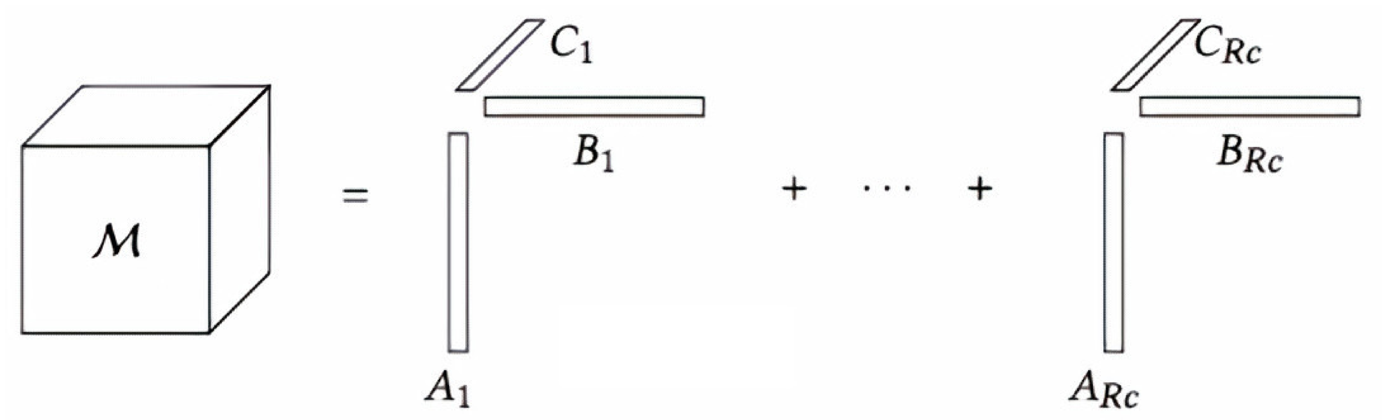
$$B = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & -1 & 0 \end{bmatrix}$$

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

$$\begin{aligned} M1 &= (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) \\ M2 &= (A_{21} + A_{22}) \cdot B_{11} \\ M3 &= A_{11} \cdot (B_{11} - B_{22}) \\ M4 &= A_{22} \cdot (B_{21} - B_{11}) \\ M5 &= (A_{11} + A_{12}) \cdot B_{22} \\ M6 &= (A_{21} - A_{11}) \cdot (B_{11} + B_{12}) \\ M7 &= (A_{12} - A_{22}) \cdot (B_{21} + B_{22}) \end{aligned}$$

$$\begin{aligned} C_{11} &= M1 + M4 - M5 + M7 \\ C_{12} &= M3 + M5 \\ C_{21} &= M2 + M4 \\ C_{22} &= M1 - M2 + M3 + M6 \end{aligned}$$

Classic CPDGN Decomposition



$$\mathcal{M} = \sum_{r=1}^{R} A_r \circ B_r \circ C_r$$

$$min\ f(A, B, C) = \frac{1}{2}\|\mathcal{M} - [A, B, C]\|^2$$

Gradient:

$$\nabla \mathbf{f} = [\mathrm{vec}(\frac{\partial f}{\partial A})^{\mathsf{T}}\ \mathrm{vec}(\frac{\partial f}{\partial B})^{\mathsf{T}}\ \mathrm{vec}(\frac{\partial f}{\partial C})^{\mathsf{T}}]^{\mathsf{T}}$$

where

$$\frac{\partial f}{\partial A} = -M_{(1)}(C \odot B) + A(C^{\mathsf{T}}C * B^{\mathsf{T}}B)$$

Jacobian:

$$\mathbf{J} = [J_A\ J_B\ J_C]$$

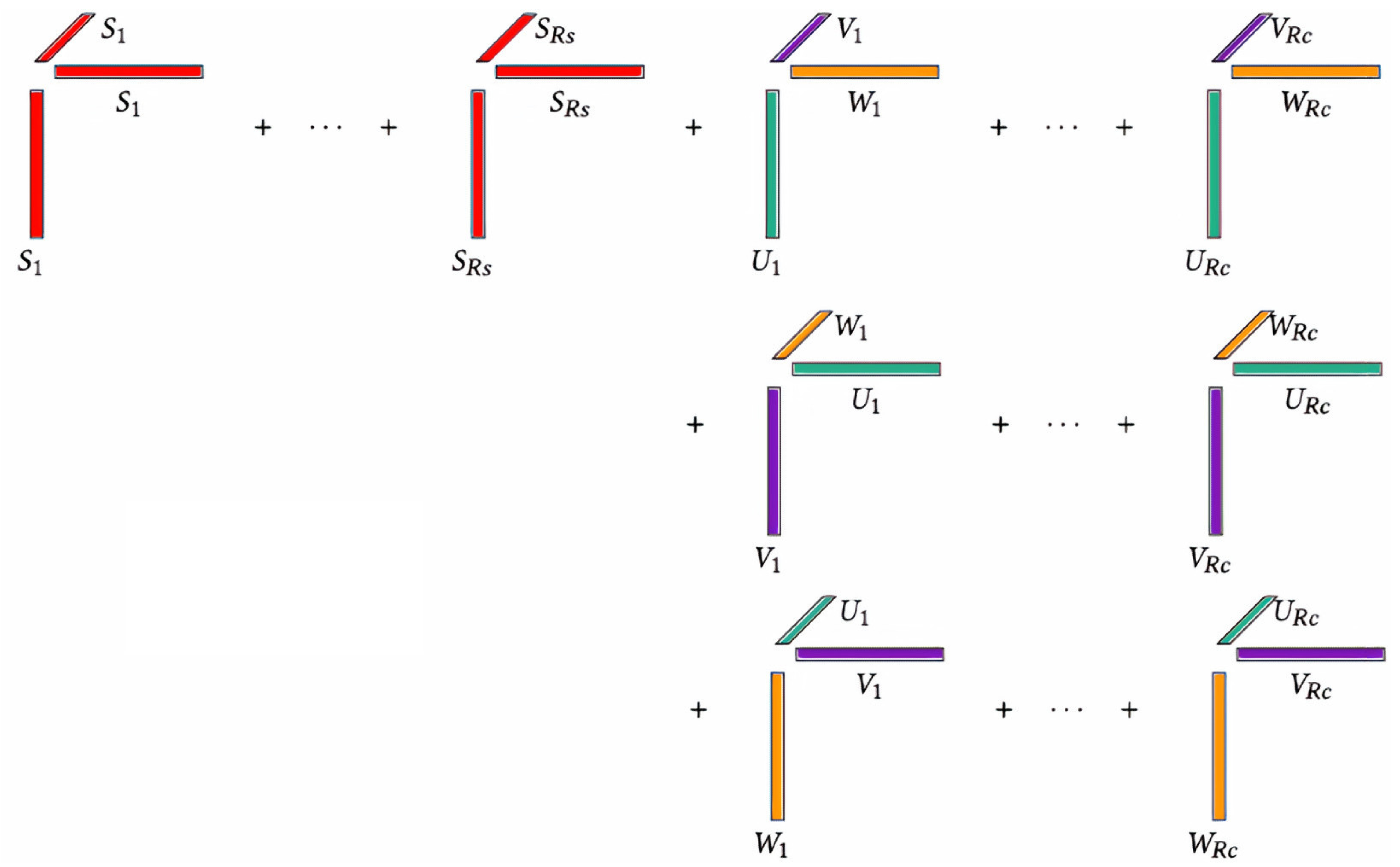
where

$$J_A = -(C \odot B) \otimes I_n$$

Modifying CPDGN to search for Cyclic Invariance

Substitutions and Derivations:

$$\mathbf{A} = [\mathbf{S\ U\ V\ W}]$$
$$\mathbf{B} = [\mathbf{S\ W\ U\ V}]$$
$$\mathbf{C} = [\mathbf{S\ V\ W\ U}]$$



$$\mathcal{M} = \sum_{q=1}^{Rs} S_q \circ S_q \circ S_q + \sum_{k=1}^{Rc}(U_k \circ V_k \circ W_k + W_k \circ U_k \circ V_k + V_k \circ W_k \circ U_k)$$

$$min\ f(S, U, V, W) = \frac{1}{2}\|\mathcal{M} - [S, S, S] - [U, V, W] - [W, U, V] - [V, W, U]\|^2$$

Gradient:

$$\nabla \mathbf{f} = [\mathrm{vec}(\frac{\partial f}{\partial S})^{\mathsf{T}}\ \mathrm{vec}(\frac{\partial f}{\partial U})^{\mathsf{T}}\ \mathrm{vec}(\frac{\partial f}{\partial V})^{\mathsf{T}}\ \mathrm{vec}(\frac{\partial f}{\partial W})^{\mathsf{T}}]^{\mathsf{T}}$$

$$\begin{aligned} \frac{\partial f}{\partial U} = \ & 3\Big(S\big((S^{\mathsf{T}}V) * (S^{\mathsf{T}}W)\big) + U\big((V^{\mathsf{T}}V) * (W^{\mathsf{T}}W)\big) \\ & + V\big((W^{\mathsf{T}}V) * (U^{\mathsf{T}}W)\big) + W\big((U^{\mathsf{T}}V) * (V^{\mathsf{T}}W)\big)\Big) \\ & - M_{(1)}(V \odot W) - M_{(2)}(W \odot V) - M_{(3)}(V \odot W) \end{aligned}$$

Jacobian:

$$\mathbf{J} = [J_s\ J_u\ J_v\ J_w] \in \mathbb{R}^{n^3 \times n(R_s + 3R_c)}$$

$$J_u = (V \odot W) \otimes I_n + \Pi_2^{\mathsf{T}} \cdot (W \odot V) \otimes I_n + \Pi_3^{\mathsf{T}} \cdot (V \odot W) \otimes I_n$$

$$J^{\mathsf{T}}J \cdot vec(G) = \begin{bmatrix} J_s^{\mathsf{T}}J_s & J_s^{\mathsf{T}}J_u & J_s^{\mathsf{T}}J_v & J_s^{\mathsf{T}}J_w \\ J_u^{\mathsf{T}}J_s & J_u^{\mathsf{T}}J_u & J_u^{\mathsf{T}}J_v & J_u^{\mathsf{T}}J_w \\ J_v^{\mathsf{T}}J_s & J_v^{\mathsf{T}}J_u & J_v^{\mathsf{T}}J_v & J_v^{\mathsf{T}}J_w \\ J_w^{\mathsf{T}}J_s & J_w^{\mathsf{T}}J_u & J_w^{\mathsf{T}}J_v & J_w^{\mathsf{T}}J_w \end{bmatrix} \begin{bmatrix} vec(G_s) \\ vec(G_u) \\ vec(G_v) \\ vec(G_w) \end{bmatrix}$$

$$J_u^{\mathsf{T}}J_s vec(G_s) = 3vec\Big(G_s(S^{\mathsf{T}}V) * (S^{\mathsf{T}}W) + S\big((G_s^{\mathsf{T}}W) * (S^{\mathsf{T}}V) + (G_s^{\mathsf{T}}V) * (S^{\mathsf{T}}W)\big)\Big)$$

Preliminary Results

Findings so far:

$\mathcal{M}_2$ Rank 7	$\mathcal{M}_3$ Rank 23	$\mathcal{M}_4$ Rank 49
Rs=4, Rc=1	Rs=11, Rc=4	Rs=16, Rc=11
Rs=1, Rc=2	Rs=5, Rc=6	Rs=1, Rc=16
	Rs=2, Rc=7	

Table: Exact Solutions through Cyclic Invariant CPDGN

There is also evidence for numerical solution in Rs=8, Rc=5 for $\mathcal{M}_3$ Rank 23 as well lower ranks. Additionally, there is evidence for several different Rs's for $\mathcal{M}_4$ Rank 49.

Future Work: Explore the possibility of numeric solutions in places named above, as well as searching for lower ranks in $\mathcal{M}_4$ and bigger Matrix Multiplication Tensors.

References

1. Rouse, Kathryn Z., and Grey M. Ballard. "On the Efficiency of Algorithms for Tensor Decompositions and Their Applications." Wake Forest University, 2018. Print.
2. G. Ballard, C. Ikenmeyer, J. Landsberg and N. Ryder, The geometry of rank decompositions of matrix multiplication II: 3x3 matrices, Journal of Pure and Applied Algebra, Volume 223, Number 8, pp. 3205 - 3224, 2018.