Optimal ordering of tensor operations for minimal flops

Raphaël Widdershoven, Lieven De Lathauwer

April 2025

1 Minimal flops for tensor-matrix products

1.1 Introduction

Given an order-N tensor of size $I_1 \times I_2 \times \cdots \times I_N$, we want to multiply it sequentially along multiple modes with matrices of size $R_n \times I_n$. The goal is to determine the optimal ordering of these multiplications to minimize the total number of floating-point operations (FLOPs). We propose that sorting the modes descending based on the value:

$$S_n = \frac{I_n - R_n}{I_n R_n} \tag{1}$$

produces an optimal ordering.

1.2 Flop Cost Analysis

Multiplying an $I_1 \times \cdots \times I_N$ tensor in mode n with a matrix of size $R_n \times I_n$ requires:

$$O\left(R_n \prod_{i=1}^{N} I_i\right) \text{ flops.} \tag{2}$$

Now, consider multiplying this tensor along two modes, n and m, using matrices $R_n \times I_n$ and $R_m \times I_m$. The total costs for the two possible orders are:

$$\text{First in mode } n \text{ then mode } m : O\left(R_n \prod_{i=1}^{N} I_i + R_n R_m \prod_{i=1, i\neq n}^{N} I_i\right). \tag{3}$$

$$\text{First in mode } m \text{ then mode } n : O\left(R_m \prod_{i=1}^{N} I_i + R_n R_m \prod_{i=1, i\neq m}^{N} I_i\right). \tag{4}$$

To determine the best order, we compare:

$$R_n \prod_{i=1}^{N} I_i + R_n R_m \prod_{i=1, i\neq n}^{N} I_i \quad \overset{?}{>} \quad R_m \prod_{i=1}^{N} I_i + R_n R_m \prod_{i=1, i\neq m}^{N} I_i. \tag{5}$$

Dividing by $R_n R_m \prod_{i=1}^{N} I_i$ yields:

$$\frac{1}{R_m} + \frac{1}{I_n} \quad \overset{?}{>} \quad \frac{1}{R_n} + \frac{1}{I_m}. \tag{6}$$

Rearranging:

$$\frac{1}{R_m} - \frac{1}{I_m} \quad \overset{?}{>} \quad \frac{1}{R_n} - \frac{1}{I_n}. \tag{7}$$

Thus, the correct ordering is obtained by sorting the values $S_n = \frac{I_n - R_n}{I_n R_n} = \frac{1}{R_n} - \frac{1}{I_n}$ in decreasing order.

1.3 Global Optimality of the Ordering

We now prove that this sorting results in a globally optimal ordering of tensor-matrix multiplications.

1. **Local Optimality:** The derived sorting criterion ensures that swapping any two adjacent operations in the sequence would increase the total number of FLOPs, meaning our ordering is a local minimum.

2. **Uniqueness of Local Minimum:** If all values S_n are distinct, the ordering is uniquely determined, meaning there is only one local minimum. Any other ordering would have two consecutive values that are decreasing. These can then be swapped for a better ordering. This means that the only local minimum is given by sorting in ascending order, which must thus be the global minimum.

3. **Equivalence of Ties:** If some values S_n are equal, swapping the corresponding modes does not change the total FLOP count, meaning all such orderings are equally optimal.

Since our method guarantees a unique or equivalently optimal local minimum, it follows that the sorted order is also globally optimal.

1.4 Conclusion

We have proven that sorting tensor-matrix products by decreasing values of $S_n = \frac{I_n - R_n}{I_n R_n}$ guarantees an optimal ordering in terms of minimizing floating-point operations. This approach provides an efficient strategy for computing multi-mode tensor-matrix products efficiently.

2 Minimal flops for sequentially truncated multi-linear singular value decomposition

In a similar fashion, we can determine the order of computing a ST-MLSVD. We make the assumption that the truncation rank is given a priori.

The first thing to determine is the cost of computing a rank-R truncated SVD of a fat $N \times M$ matrix:

Full spectrum SVD	$\sim 4MN^2 - \frac{4}{3}N^3$
Lanczos/randomized SVD	$\sim 4RMN + 6R^3$
+ product $\mathbf{\Sigma V}^{\mathrm{T}}$	$\sim RM$

Note that in ST-MLSVD, for a tensor of size $I_1 \times \cdots \times I_N$ truncated in mode n to rank-R_n, $R = R_n$, $N = I_n$ and $M = \prod_{i=1, i \neq n}^{N} I_i$.

We will work out four formulas, based on assuming different total costs:

Cost A	$\sim 4MN^2$	section 2.1
Cost B	$\sim 4MN^2 + RM$	section 2.2
Cost C	$\sim 4RMN$	section 1
Cost D	$\sim 4RMN + RM$	section 2.3

Note that the cube terms ($-\frac{4}{3}N^3$ and $6R^3$) are left out of these costs, as $R < N \ll M$ and otherwise the equations would not simplify easily. Cost C is (up to a constant factor) the same as for tmprod, so the derivation can be found there.

Note also that the term RM, although it does involve the large dimension M, is typically (significantly) smaller than $4MN^2$ and $4RMN$.

An overview of the obtained formulas to sort on is given in section 2.4.

2.1 A: Full spectrum SVD

We consider an expression for the cost of $O(MN^2)$. Consider the ordering of two modes, n and m that need to be truncated sequentially, with total costs:

$$\text{Mode } n \text{ then } m : O\left(I_n \prod_{i=1}^{N} I_i + I_m R_n \prod_{i=1, i\neq n}^{N} I_i\right). \tag{8}$$

$$\text{Mode } m \text{ then } n : O\left(I_m \prod_{i=1}^{N} I_i + I_n R_m \prod_{i=1, i\neq m}^{N} I_i\right). \tag{9}$$

To determine the best order, we compare:

$$I_n \prod_{i=1}^{N} I_i + I_m R_n \prod_{i=1, i\neq n}^{N} I_i \quad \overset{?}{>} \quad I_m \prod_{i=1}^{N} I_i + I_n R_m \prod_{i=1, i\neq m}^{N} I_i, \tag{10}$$

$$\Downarrow / \left(I_n I_m \prod_{i=1}^{N} I_i\right) \tag{11}$$

$$\frac{1}{I_m} + \frac{R_n}{I_n^2} \quad \overset{?}{>} \quad \frac{1}{I_n} + \frac{R_m}{I_m^2}, \tag{12}$$

$$\Downarrow -\frac{R_n}{I_n^2} - \frac{R_m}{I_m^2} \tag{13}$$

$$\frac{1}{I_m} - \frac{R_m}{I_m^2} \quad \overset{?}{>} \quad \frac{1}{I_n} - \frac{R_n}{I_n^2}. \tag{14}$$

Following a similar reasoning and proof as in section 1, we conclude that sorting descendingly on

$$S_n = \frac{I_n - R_n}{I_n^2} = \frac{1}{I_n} - \frac{R_n}{I_n^2} \tag{15}$$

yields the optimal ordering.

2.2 B: Full spectrum SVD + product

We consider an expression for the cost of $O(MN^2 + \alpha RM)$ with $\alpha = \frac{1}{4}$. Consider the ordering of two modes, n and m that need to be truncated sequentially, with total costs:

$$\text{Mode } n \text{ then } m : O\left((I_n^2 I_m + \alpha R_n I_m + R_n I_m^2 + \alpha R_n R_m) \prod_{i=1, i\neq n, i\neq m}^{N} I_i\right). \tag{16}$$

$$\text{Mode } m \text{ then } n : O\left((I_n I_m^2 + \alpha R_m I_n + R_m I_n^2 + \alpha R_n R_m) \prod_{i=1, i\neq n, i\neq m}^{N} I_i\right). \tag{17}$$

To determine the best order, we divide by $\prod_{i=1}^{N} I_i$ and compare:

$$I_n + \alpha\frac{R_n}{I_n} + I_m\frac{R_n}{I_n} + \alpha\frac{R_nR_m}{I_nI_m} \quad \overset{?}{>} \quad I_m + \alpha\frac{R_m}{I_m} + I_n\frac{R_m}{I_m} + \alpha\frac{R_nR_m}{I_nI_m}, \tag{18}$$

$$\Downarrow -\alpha\frac{R_nR_m}{I_nI_m} \tag{19}$$

$$(I_m+\alpha)\left(\frac{I_n}{\alpha}+\frac{R_n}{I_n}\right) - \frac{I_nI_m}{\alpha} \quad \overset{?}{>} \quad (I_n+\alpha)\left(\frac{I_m}{\alpha}+\frac{R_m}{I_m}\right) - \frac{I_nI_m}{\alpha}, \tag{20}$$

$$\Downarrow +\frac{I_nI_m}{\alpha} \text{ and then } \Big/\left(\left(\frac{I_n}{\alpha}+\frac{R_n}{I_n}\right)\left(\frac{I_m}{\alpha}+\frac{R_m}{I_m}\right)\right) \tag{21}$$

$$\frac{(I_m+\alpha)}{\left(\frac{I_m}{\alpha}+\frac{R_m}{I_m}\right)} \quad \overset{?}{>} \quad \frac{(I_n+\alpha)}{\left(\frac{I_n}{\alpha}+\frac{R_n}{I_n}\right)}. \tag{22}$$

Following a similar reasoning and proof as in section 1, we conclude that sorting descendingly on

$$S_n = \frac{\alpha^2 I_n + \alpha I_n^2}{I_n^2 + \alpha R_n} = \frac{(I_n+\alpha)}{\left(\frac{I_n}{\alpha}+\frac{R_n}{I_n}\right)} \tag{23}$$

yields the optimal ordering.

2.3 D: Lanczos/randomized + product

We consider an expression for the cost of $O(RMN+\alpha RM)$ with $\alpha = \frac{1}{4}$. Consider the ordering of two modes, n and m that need to be truncated sequentially, with total costs:

$$\text{Mode } n \text{ then } m : O\left((R_nI_nI_m + \alpha R_nI_m + R_nR_mI_m + \alpha R_nR_m)\prod_{i=1,i\neq n,i\neq m}^{N} I_i\right). \tag{24}$$

$$\text{Mode } m \text{ then } n : O\left((R_mI_nI_m + \alpha R_mI_n + R_nR_mI_n + \alpha R_nR_m)\prod_{i=1,i\neq n,i\neq m}^{N} I_i\right). \tag{25}$$

To determine the best order, we divide by $R_nR_m\prod_{i=1}^{N} I_i$ and compare:

$$\frac{1}{R_m} + \alpha\frac{1}{R_mI_n} + \frac{1}{I_n} + \alpha\frac{1}{I_nI_m} \quad \overset{?}{>} \quad \frac{1}{R_n} + \alpha\frac{1}{R_nI_m} + \frac{1}{I_m} + \alpha\frac{1}{I_nI_m}, \tag{26}$$

$$\Downarrow -\alpha\frac{1}{I_nI_m} \tag{27}$$

$$\alpha\left(\frac{1}{I_n}+\frac{1}{\alpha}\right)\left(\frac{1}{R_m}+\frac{1}{\alpha}\right) - \frac{1}{\alpha} \quad \overset{?}{>} \quad \alpha\left(\frac{1}{I_m}+\frac{1}{\alpha}\right)\left(\frac{1}{R_n}+\frac{1}{\alpha}\right) - \frac{1}{\alpha}, \tag{28}$$

$$\Downarrow +\frac{1}{\alpha} \text{ and then } \Big/\left(\alpha\left(\frac{1}{I_n}+\frac{1}{\alpha}\right)\left(\frac{1}{I_m}+\frac{1}{\alpha}\right)\right) \tag{29}$$

$$\frac{\left(\frac{1}{R_m}+\frac{1}{\alpha}\right)}{\left(\frac{1}{I_m}+\frac{1}{\alpha}\right)} \quad \overset{?}{>} \quad \frac{\left(\frac{1}{R_n}+\frac{1}{\alpha}\right)}{\left(\frac{1}{I_n}+\frac{1}{\alpha}\right)}. \tag{30}$$

Following a similar reasoning and proof as in section 1, we conclude that sorting descendingly on

$$S_n = \frac{I_nR_n + \alpha I_n}{I_nR_n + \alpha R_n} = \frac{\left(\frac{1}{R_n}+\frac{1}{\alpha}\right)}{\left(\frac{1}{I_n}+\frac{1}{\alpha}\right)} \tag{31}$$

yields the optimal ordering.

2.4 Overview results

The optimal order is obtained by sorting descendingly on S_n, given the following presumed form of cost function (with $\alpha = \frac{1}{4}$):

Cost	S_n
$\sim 4MN^2$	$\frac{I_n - R_n}{I_n^2}$
$\sim 4MN^2 + RM$	$\frac{\alpha^2 I_n + \alpha I_n^2}{I_n^2 + \alpha R_n}$
$\sim 4RMN$	$\frac{I_n - R_n}{I_n R_n}$
$\sim 4RMN + RM$	$\frac{I_n R_n + \alpha I_n}{I_n R_n + \alpha R_n}$