Source task and conventions

Compare $(AB)C$ with $A(BC)$ for shapes $2^{10} \times 2^{16}$, $2^{16} \times 2^{5}$, and $2^{5} \times 2^{14}$.

Conventional solution and proof

For dense classical multiplication of an $r \times s$ matrix by an $s \times t$ matrix, each of the rt outputs is a length-s dot product: rst multiplications and $rt(s-1)$ additions. Thus $(AB)C$ uses

$$2^{31} + 2^{29} = 2,684,354,560$$

multiplications. The alternative $A(BC)$ uses

$$2^{35} + 2^{40} = 1,133,871,366,144.$$

The exact multiplication-count ratio is 422.4. The left intermediate has 2^{15} entries; the right has 2^{30}. At eight bytes per entry these occupy 256 KiB and 8 GiB, respectively: a factor of $32,768$. Common input and final-output storage is excluded from both intermediate counts. Associativity over exact real arithmetic proves output equivalence; floating-point association can give different rounding errors.

Structural or analytic derivation

Treat matrix multiplication as contraction of a three-edge tensor network. Eliminating the large 2^{16} index first produces the small $2^{10} \times 2^{5}$ intermediate. Enumerating the two possible binary association trees proves that the left association minimizes classical multiplication count and intermediate size among these two schedules. The random Gaussian initialization does not change these dimension-based counts.