Fast Linear Algebra

Strassen’s Matrix Multiplication

\[ A, B \in \mathbb{R}^{n \times n}, \text{ ring.} \]

Traditional Matrix Multiplication

\[ \rightarrow C = A \cdot B \text{ - row by column inner product} \]

\[ \rightarrow n^2 \text{ - inner products} \]

\[ \rightarrow \text{Each are } n \text{ multiplication and } n - 1 \text{ additions. } n^3 - n^2 \text{ operations in } R \]

Recursive Matrix Multiplication

Assume \( n = 2^k \), \( A, B \in \mathbb{R}^n \).

\[
A = \begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{pmatrix}
B = \begin{pmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{pmatrix}
\]

\[ A_{ij} \cdot B_{kl} \in \mathbb{R}^{\frac{n}{2} \times \frac{n}{2}} \]

Regular 2 \times 2 matrix multiplication takes

\[ \cdot \text{ 8 multiplications} \]

\[ \cdot \text{ 4 additions} \]

Recursion:

\[ \cdot \text{ 8 multiplications on } \frac{n}{2} \times \frac{n}{2} \text{ matrices.} \]

\[
C = \begin{pmatrix}
A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\
A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22}
\end{pmatrix}
\]

Regular Matrix Multiplication Complexity:

\[
M(n) = 8M\left(\frac{n}{2}\right) + 4\left(\frac{n}{2}\right)^2
\]

\[ \Rightarrow M(n) \in \Theta(n^3) \]

With only 7 multiplications and \( c \) additions:

\[
M(n) = 7M\left(\frac{n}{2}\right) + c\left(\frac{n}{2}\right)^2
\]

\[ \Rightarrow M(n) \in \Theta(n^{\log_2 7}) \approx \Theta(n^{2.807}) \]
Strassen's Algorithm with 7 multiplications for $2 \times 2$ matrices

A, B as above. Then $C = AB$ can be computed as follow:

\begin{align*}
P_1 &= (A_{11} + A_{22})(B_{11} + B_{22}) \\
P_2 &= (A_{21} + A_{22})B_{11} \\
P_3 &= A_{11}(B_{12} - B_{22}) \\
P_4 &= A_{22}(-B_{11} + B_{21}) \\
P_5 &= (A_{11} + A_{12})B_{22} \\
P_6 &= (-A_{11} + A_{21})(B_{11} + B_{12}) \\
P_7 &= (A_{12} - A_{22})(B_{21} + B_{22})
\end{align*}

\begin{align*}
C_{11} &= P_1 + P_4 - P_5 + P_7 \\
C_{21} &= P_2 + P_4 \\
C_{12} &= P_3 + P_5 \\
C_{22} &= P_1 + P_3 - P_2 + P_6
\end{align*}

Below is a different formula also with 7 multiplications

\begin{align*}
S_1 &= A_{21} + A_{22} \\
S_2 &= S_1 - A_{11} \\
S_3 &= A_{11} - A_{21} \\
S_4 &= A_{12} - S_2 \\
S_5 &= S_1 - A_{11} \\
S_6 &= S_2 - T_1 \\
T_1 &= B_{12} - B_{11} \\
T_2 &= B_{22} - B_{21} \\
T_3 &= B_{22} - B_{21} \\
T_4 &= T_2 - B_{21}
\end{align*}

\begin{align*}
P_1 &= A_{11}B_{11} \\
P_2 &= A_{12}B_{21} \\
P_3 &= S_4B_{22} \\
P_4 &= A_{22}T_4 \\
P_5 &= S_1T_1 \\
P_6 &= S_2T_2 \\
P_7 &= S_3T_3
\end{align*}

\begin{align*}
C_{11} &= U_1 = P_1 + P_2 \\
C_{12} &= U_5 = U_4 + P_3 \\
C_{21} &= U_6 = U_3 - P_4 \\
C_{22} &= U_7 = U_3 + P_5
\end{align*}

**Definition 1.** Exponent $\omega$ of matrix multiplication:

$$\omega = \inf \{ h \in \mathbb{R} : 2 \times 2 \text{ matrices over } \mathbb{R} \text{ can be multiplied using } O(n^h) \text{ operations in } \mathbb{R} \}$$

$$\Rightarrow \omega \leq \log_2 7$$

**Matrix Multiplication and Tensors**

**Matrix multiplication as bilinear map**

$U, V, W$ are vector spaces over a field $F$, and consider a map $f : U \times V \rightarrow W$. Then $f$ is **bilinear** if it satisfies the following properties:

- $f(\alpha u + \beta u', v) = \alpha f(u, v) + \beta f(u', v)$,
- $f(u, \alpha v + \beta v') = \alpha f(u, v) + \beta f(u, v')$ for $u, u' \in U$, $v, v' \in V$.  

2
Matrix Multiplication

Define the bilinear map

\[ \mathcal{M}_{n \times n} : \mathbb{F}^{n \times n} \times \mathbb{F}^{n \times n} \to \mathbb{F}^{n \times n} \]

\[ (A, B) \mapsto C = A \cdot B \]

e.g., \[ \mathcal{M}_{2 \times 2} : \mathbb{F}^{2 \times 2} \times \mathbb{F}^{2 \times 2} \to \mathbb{F}^{2 \times 2} \]

A bilinear map \( \phi : \mathbb{F}^m \times \mathbb{F}^m \to \mathbb{F} \) can be described by a matrix once we fix a basis \( \{ e_1, \ldots, e_m \} \subset \mathbb{F}^m \):

\[ M = [\phi(e_i, e_j)]_{i,j=1}^m, \quad \phi(u, v) = u^T M v \]

Conversely every matrix corresponds to a bilinear product.

\[ A \in \mathbb{F}^{m \times m}, \quad \phi : \mathbb{F}^m \times \mathbb{F}^m \to \mathbb{F} \]

\[ (u, v) \mapsto u^T Av, \text{ bilinear.} \]

Here we have

\[ \mathcal{M}_{2 \times 2} : \mathbb{F}^{2 \times 2} \times \mathbb{F}^{2 \times 2} \to \mathbb{F}^{2 \times 2} \]

\[ \uparrow \]

\[ \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \]

Each entry is bilinear \( \Rightarrow \) 4 bilinear maps \( \mathcal{C}_{ij} : \mathbb{F}^{2 \times 2} \times \mathbb{F}^{2 \times 2} \to \mathbb{F} \).

Fix the standard basis \( \{ E_{11}, E_{12}, E_{21}, E_{22} \} \subset \mathbb{F}^{2 \times 2} \)

Then \( \mathcal{C}_{11} = A_{11}B_{11} + A_{12}B_{21} \) corresponds to a 4 \( \times \) 4 matrix:

\[ M_{11} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

Similarly for \( \mathcal{C}_{12} = A_{11}B_{12} + A_{12}B_{22} \),

\[ M_{12} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

etc, so we get 4 matrices each of size 4 \( \times \) 4, each with 2 non-zero entries. We will stack them to get a 3-dimensional tensor \( M \in \mathbb{Q}^{4 \times 4 \times 4} \) (see Figure 1 and 2).

\[ \begin{array}{c}
M_{11} \\
M_{12} \\
M_{21} \\
M_{22}
\end{array} \]

Figure 1: The 4 \( \times \) 4 \( \times \) 4 tensor \( M \) corresponding to the bilinear map \( \mathcal{M}_{2 \times 2} \)
Definition 2 (Contraction). Let \( u, v, w, a, b \in F^4 \) and \( T \in F^{4 \times 4 \times 4} \). We denote the tensor product by \( u \otimes v \otimes w := [u_i v_j w_k]_{i,j,k=1}^4 \). We denote the contraction by the first and the second indices by \( T \odot_1 a \odot_2 b := [\sum_{i,j=1}^4 T_{ijk} a_i b_j]_{k=1}^4 \in F^4 \).

Note: if \( M \) is the \( 4 \times 4 \times 4 \) tensor associated to \( \mathcal{M}_{2 \times 2} \) and

\[
\begin{bmatrix}
a_{11} \\
a_{12} \\
a_{21} \\
a_{22}
\end{bmatrix}, \quad \begin{bmatrix}
b_{11} \\
b_{12} \\
b_{21} \\
b_{22}
\end{bmatrix}
\]

then the array \( M \odot_1 a \odot_2 b \) gives the formula for the matrix multiplication:

\[
M \odot_1 a \odot_2 b = [a^T M_{11} b, a^T M_{12} b, a^T M_{21} b, a^T M_{22} b]^T = [a_{11} b_{11} + a_{12} b_{21}, a_{11} b_{12} + a_{12} b_{22}, a_{21} b_{11} + a_{22} b_{21}, a_{21} b_{12} + a_{22} b_{22}]^T.
\]

Example:

\[
u = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad a = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}, \quad a \odot u = l_u(a) = a_{11} + 2a_{12} + 3a_{21} + 4a_{22}.
\]

Definition 3 (Rank 1 Tensor). We say that tensor \( T \in F^{4 \times 4 \times 4} \) has rank-1 if \( \exists u, v, w \in F^4 \) such that \( T = u \otimes v \otimes w = [u_i v_j w_k]_{i,j,k=1}^4 \).

Example

\( e_2 \otimes e_1 \otimes e_3 \in F^{4 \times 4 \times 4} \) has only the third layer non-zero. This second layer is

\[
e_2 \otimes e_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\]

Example:
Construct the tensor product \( T = u \otimes v \otimes w = = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{F}^{2 \times 2 \times 2}. \)

Below is a view of \( T \) as per the view explored above.

![Figure 3: \( u \otimes v \otimes w \)](image)

Note: \( T \) is rank-1 if each layer of \( T \) is a rank-1 matrix and each layer is a multiple of the others.

Remark: For rank-1 tensors, contraction can be computed using 1 multiplication. If \( T = u \otimes v \otimes w \) then

\[(u \otimes v \otimes w) \odot a \odot b = u \odot a \otimes v \odot b \otimes w = l_u(a) \otimes l_v(b) \otimes w.\]

Thus, if \( T \in \mathbb{F}^{4 \times 4 \times 4} \), \( \text{rank}(T) = 1 \) then \( T \odot a \odot b \) can be computed using 1 multiplication of linear combinations of elements in \( a \) and \( b \) and 6 additions and 8 scalar multiplications.

**Definition 4** (Rank of Tensor). \( T \in \mathbb{F}^{4 \times 4 \times 4} \) has rank \( r \) if \( r \) is minimal such that \( T = T_1 + \cdots + T_r \) and \( T_i \) are rank-1.

Example: We show how to write the \( 4 \times 4 \times 4 \) tensor \( M \) as a sum of 8 rank one tensors, corresponding to the eight non-zero entries of that tensor.

\[
M = e_1 \otimes e_1 \otimes e_1 + e_2 \otimes e_3 \otimes e_1 + e_1 \otimes e_2 \otimes e_2 + e_2 \otimes e_3 \otimes e_2 + e_3 \otimes e_1 \otimes e_3 + e_4 \otimes e_3 \otimes e_4 + e_2 \otimes e_3 \otimes e_4.
\]

Thus \( \text{Rank}(M) \leq 8 \). Below, we will prove that \( \text{Rank}(M) \leq 7 \), using Strassen's formula.

Remark: If \( \text{Rank}(T) = r \) then contraction can be computed using \( r \) multiplication.

**Strassen's Algorithm as the rank-7 decomposition of \( M \)**

\( P_1, \ldots, P_7 \) correspond to rank-1 \( 4 \times 4 \) matrices.

Example:

\[
P_1 = (A_{11} + A_{22})(B_{11} + B_{22}) \Rightarrow u_1 \otimes v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}
\]

\[
P_2 = (A_{21} + A_{22})B_{11} \Rightarrow u_2 \otimes v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}
\]

\[
\text{etc.}
\]

Then

\[
M = u_1 \otimes v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + u_2 \otimes v_2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + u_3 \otimes v_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + u_4 \otimes v_4 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + u_5 \otimes v_5 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + u_6 \otimes v_6 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + u_7 \otimes v_7 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}
\]
⇒ \( \text{rank}(M) \leq 7. \)