

Fast Matrix Multiplication

Michael Schmid

31.05.2021

Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

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

Strassen's Algorithm

$$\mathbf{AB} = \mathbf{C}$$

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

$$C_{11} = A_{11} \cdot B_{11} + A_{12} \cdot B_{21}$$

$$C_{12} = A_{11} \cdot B_{12} + A_{12} \cdot B_{22}$$

$$C_{21} = A_{21} \cdot B_{11} + A_{22} \cdot B_{21}$$

$$C_{22} = A_{21} \cdot B_{12} + A_{22} \cdot B_{22}$$

Algorithm 1 Square Matrix Multiplication

```
1: function MM(A, B, C,  $n$ )
2:    $sum \leftarrow 0$ 
3:   for  $i = 0, 1, 2, \dots, n - 1$  do
4:     for  $j = 0, 1, 2, \dots, n - 1$  do
5:        $sum \leftarrow 0$ 
6:       for  $k = 0, 1, 2, \dots, n - 1$  do
7:          $sum \leftarrow sum + \mathbf{A}[i][k] \cdot \mathbf{B}[k][j]$ 
8:        $\mathbf{C}[i][j] \leftarrow sum$ 
9:   return C
```

CNN Parallel Computing

$$\begin{bmatrix} A_{1,1} & \cdots & A_{1,k} & \cdots & A_{1,n} \\ \vdots & & \vdots & & \vdots \\ A_{i,1} & \cdots & A_{i,k} & \cdots & A_{i,n} \\ \vdots & & \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,k} & \cdots & A_{m,n} \end{bmatrix} \begin{bmatrix} B_{1,1} & \cdots & B_{1,j} & \cdots & B_{1,p} \\ \vdots & & \vdots & & \vdots \\ B_{k,1} & \cdots & B_{k,j} & \cdots & B_{i,p} \\ \vdots & & \vdots & & \vdots \\ B_{n,1} & \cdots & B_{n,j} & \cdots & B_{n,p} \end{bmatrix}$$
$$\begin{bmatrix} C_{1,1} & \cdots & C_{1,j} & \cdots & C_{1,p} \\ \vdots & & \vdots & & \vdots \\ C_{i,1} & \cdots & C_{i,j} & \cdots & C_{i,p} \\ \vdots & & \vdots & & \vdots \\ C_{m,1} & \cdots & C_{m,j} & \cdots & C_{m,p} \end{bmatrix}$$

Big \mathcal{O} notation

- Time complexity of an algorithm

Big \mathcal{O} notation

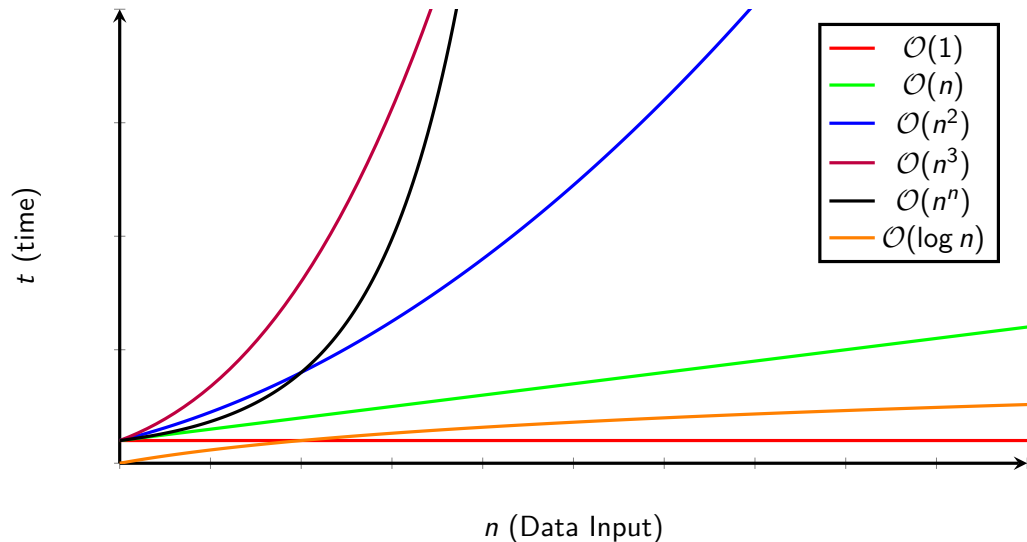
- Time complexity of an algorithm
- How many multiplications in a function

Big \mathcal{O} notation

Algorithm 2 Square Matrix Multiplication

```
1: function MM(A, B, C,  $n$ )
2:    $sum \leftarrow 0$ 
3:   for  $i = 0, 1, 2 \dots, n - 1$  do
4:     for  $j = 0, 1, 2 \dots, n - 1$  do
5:        $sum \leftarrow 0$ 
6:       for  $k = 0, 1, 2 \dots, n - 1$  do
7:          $sum \leftarrow sum + \mathbf{A}[i][k] \cdot \mathbf{B}[k][j]$ 
8:        $\mathbf{C}[i][j] \leftarrow sum$ 
9:   return C
```

Big \mathcal{O} notation



Strassen's Algorithm

$$\text{I} = (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$\text{II} = (A_{21} + A_{22}) \cdot B_{11}$$

$$\text{III} = A_{11} \cdot (B_{12} - B_{22})$$

$$\text{IV} = A_{22} \cdot (-B_{11} + B_{21})$$

$$\text{V} = (A_{11} + A_{12}) \cdot B_{22}$$

$$\text{VI} = (-A_{11} + A_{21}) \cdot (B_{11} + B_{12})$$

$$\text{VII} = (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

Strassen's Algorithm

$$I = (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$II = (A_{21} + A_{22}) \cdot B_{11}$$

$$III = A_{11} \cdot (B_{12} - B_{22})$$

$$IV = A_{22} \cdot (-B_{11} + B_{21})$$

$$V = (A_{11} + A_{12}) \cdot B_{22}$$

$$VI = (-A_{11} + A_{21}) \cdot (B_{11} + B_{12})$$

$$VII = (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{11} = I + IV - V + VII$$

$$C_{21} = II + IV$$

$$C_{12} = III + V$$

$$C_{22} = I + III - II + VI$$

Strassen's Algorithm

$$I = (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$II = (A_{21} + A_{22}) \cdot B_{11}$$

$$III = A_{11} \cdot (B_{12} - B_{22})$$

$$IV = A_{22} \cdot (-B_{11} + B_{21})$$

$$V = (A_{11} + A_{12}) \cdot B_{22}$$

$$VI = (-A_{11} + A_{21}) \cdot (B_{11} + B_{12})$$

$$VII = (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{11} = I + IV - V + VII$$

$$C_{21} = II + IV$$

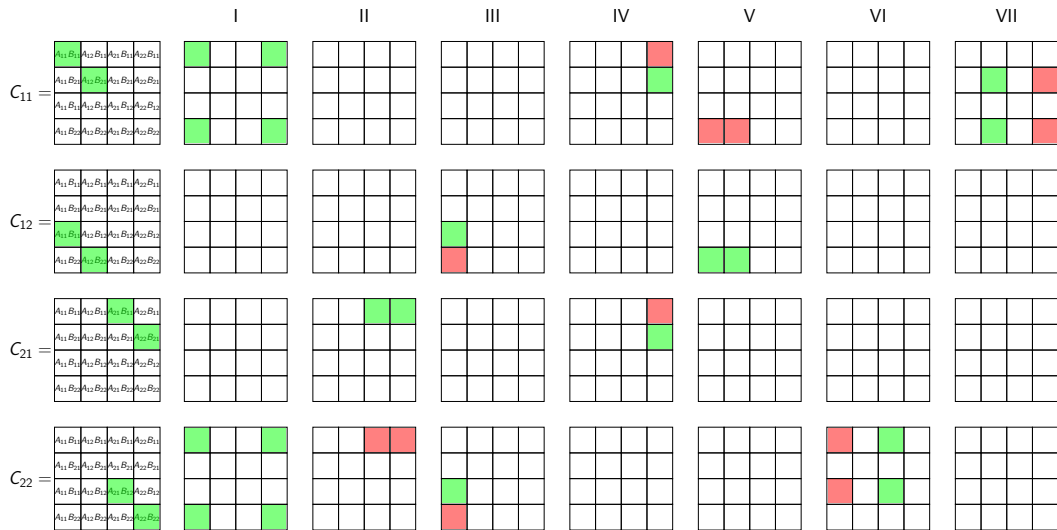
$$C_{12} = III + V$$

$$C_{22} = I + III - II + VI$$

$$C_{11} = (A_{11} + A_{22}) \cdot (B_{11} + B_{22}) + A_{22} \cdot (-B_{11} + B_{21}) - (A_{11} + A_{12}) \cdot B_{22} + (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{11} = A_{11}B_{11} + A_{11}B_{22} + A_{22}B_{11} + A_{22}B_{22} - A_{22}B_{11} + A_{22}B_{21} - A_{11}B_{22} - A_{12}B_{22} + A_{12}B_{21} + A_{12}B_{22} - A_{22}B_{21} - A_{22}B_{22}$$

$$C_{11} = A_{11}B_{11} + A_{12}B_{21}$$



Strassen's Algorithm

$$I = (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$II = (A_{21} + A_{22}) \cdot B_{11}$$

$$III = A_{11} \cdot (B_{12} - B_{22})$$

$$IV = A_{22} \cdot (-B_{11} + B_{21})$$

$$V = (A_{11} + A_{12}) \cdot B_{22}$$

$$VI = (-A_{11} + A_{21}) \cdot (B_{11} + B_{12})$$

$$VII = (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{11} = I + IV - V + VII$$

$$C_{21} = II + IV$$

$$C_{12} = III + V$$

$$C_{22} = I + III - II + VI$$

$$\mathbf{I} = (\mathbf{A}_{11} + \mathbf{A}_{22}) \cdot (\mathbf{B}_{11} + \mathbf{B}_{22})$$

$$\mathbf{II} = (\mathbf{A}_{21} + \mathbf{A}_{22}) \cdot \mathbf{B}_{11}$$

$$\mathbf{III} = \mathbf{A}_{11} \cdot (\mathbf{B}_{12} - \mathbf{B}_{22})$$

$$\mathbf{IV} = \mathbf{A}_{22} \cdot (-\mathbf{B}_{11} + \mathbf{B}_{21})$$

$$\mathbf{V} = (\mathbf{A}_{11} + \mathbf{A}_{12}) \cdot \mathbf{B}_{22}$$

$$\mathbf{VI} = (-\mathbf{A}_{11} + \mathbf{A}_{21}) \cdot (\mathbf{B}_{11} + \mathbf{B}_{12})$$

$$\mathbf{VII} = (\mathbf{A}_{12} - \mathbf{A}_{22}) \cdot (\mathbf{B}_{21} + \mathbf{B}_{22})$$

$$\mathbf{C}_{11} = \mathbf{I} + \mathbf{IV} - \mathbf{V} + \mathbf{VII}$$

$$\mathbf{C}_{21} = \mathbf{II} + \mathbf{IV}$$

$$\mathbf{C}_{12} = \mathbf{III} + \mathbf{V}$$

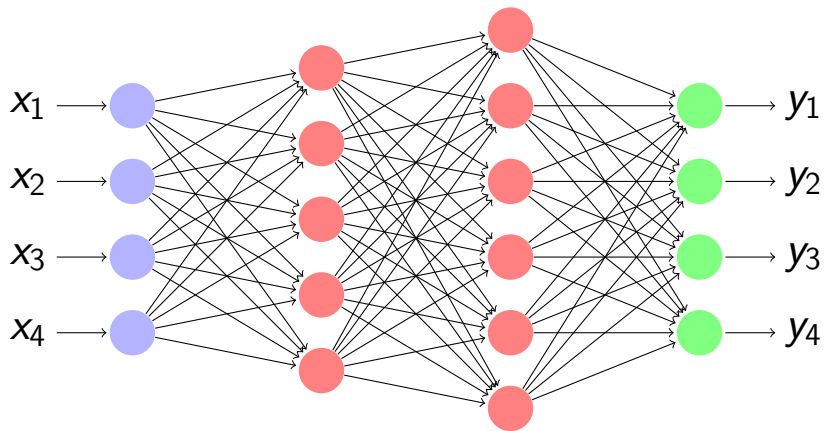
$$\mathbf{C}_{22} = \mathbf{I} + \mathbf{III} - \mathbf{II} + \mathbf{VI}$$

Algorithm

Algorithm 3 Strassen Matrix Multiplication

```
1: function MM(A, B, C,  $n$ )
2:   if  $n = 2$  then
3:      $C \leftarrow \text{zeros}((n, n))$ 
4:      $P \leftarrow (A[0][0] + A[1][1]) \cdot (B[0][0] + B[1][1])$ 
5:      $Q \leftarrow (A[1][0] + A[1][1]) \cdot B[0][0]$ 
6:      $R \leftarrow A[0][0] \cdot (B[0][1] - B[1][1])$ 
7:      $S \leftarrow A[1][1] \cdot (B[1][0] - B[0][0])$ 
8:      $T \leftarrow (A[0][0] + A[0][1]) \cdot B[1][1]$ 
9:      $U \leftarrow (A[1][0] - A[0][0]) \cdot (B[0][0] + B[0][1])$ 
10:     $V \leftarrow (A[0][1] - A[1][1]) \cdot (B[1][0] + B[1][1])$ 
11:     $C[0][0] \leftarrow P + S - T + V$ 
12:     $C[0][1] \leftarrow R + T$ 
13:     $C[1][0] \leftarrow Q + S$ 
14:     $C[1][1] \leftarrow P + R - Q + U$ 
15:  else
16:     $m \leftarrow n/2$ 
17:     $A11, A12, A21, A22 \leftarrow A[:m][:m], A[:m][m:], A[m:][:m], A[m:][:m:]$ 
18:     $B11, B12, B21, B22 \leftarrow B[:m][:m], B[:m][m:], B[m:][:m], B[m:][:m:]$ 
19:     $C11 \leftarrow \text{strassen}(A11, B11, m) + \text{strassen}(A12, B21, m)$ 
20:     $C12 \leftarrow \text{strassen}(A11, B12, m) + \text{strassen}(A12, B22, m)$ 
21:     $C21 \leftarrow \text{strassen}(A21, B11, m) + \text{strassen}(A22, B21, m)$ 
22:     $C22 \leftarrow \text{strassen}(A21, B12, m) + \text{strassen}(A22, B22, m)$ 
23:     $C \leftarrow \text{vstack}((\text{hstack}((C11, C12)), \text{hstack}((C21, C22))))$ 
24:  return C
```

Neural Network



Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix} \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

The diagram illustrates the calculation of the element $C_{1,1}$ in the resulting matrix C . It shows the first row of matrix A (containing $A_{1,1}, A_{1,2}, A_{1,3}, A_{1,4}$) and the first column of matrix B (containing $B_{1,1}, B_{2,1}, B_{3,1}, B_{4,1}$). Curved lines connect each element of the first row of A to the corresponding element in the first column of B , representing the dot product. The result of this operation is the element $C_{1,1}$ in the first row and first column of matrix C .

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \cdot \begin{bmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ B_{4,1} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

The diagram illustrates the calculation of the first row of the result matrix C from the first row of matrix A and matrix B . Matrix A is a 1x4 row vector, and matrix B is a 4x5 matrix. The first row of C is a 1x5 row vector. Curved lines connect each element of the first row of A to the corresponding column of B , showing that the first row of C is the dot product of the first row of A with the columns of B .

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix} \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

Vector-Matrix Multiplication

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & B_{1,4} & B_{1,5} \\ B_{2,1} & B_{2,2} & B_{2,3} & B_{2,4} & B_{2,5} \\ B_{3,1} & B_{3,2} & B_{3,3} & B_{3,4} & B_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & B_{4,4} & B_{4,5} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \end{bmatrix}$$

The diagram illustrates the calculation of the first element of the resulting vector C , $C_{1,1}$. It shows the dot product of the first row of matrix A (containing $A_{1,1}, A_{1,2}, A_{1,3}, A_{1,4}$) and the third row of matrix B (containing $B_{3,1}, B_{3,2}, B_{3,3}, B_{3,4}, B_{3,5}$). Curved lines connect each element of the first row of A to its corresponding element in the third row of B , representing the multiplication of pairs $(A_{1,i}, B_{3,i})$ for $i=1, 2, 3, 4$. These products are then summed to yield $C_{1,1}$.

