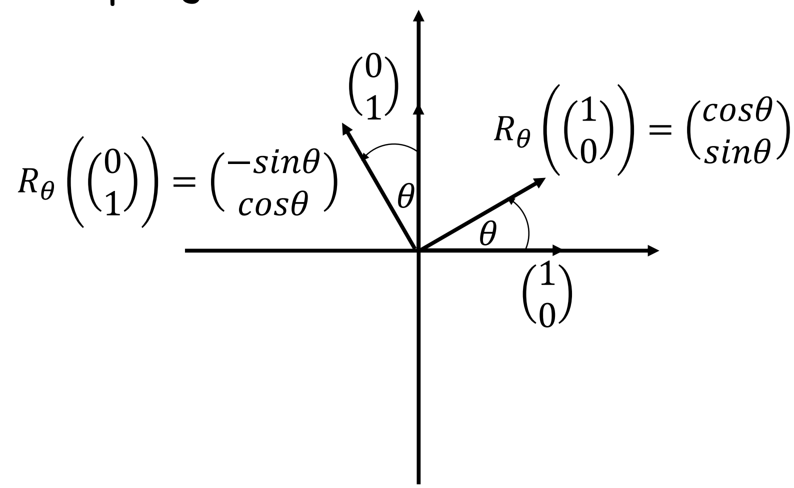
Matrix-Matrix Multiplication

Jirasak Sittigorn
Department of Computer Engineering
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang

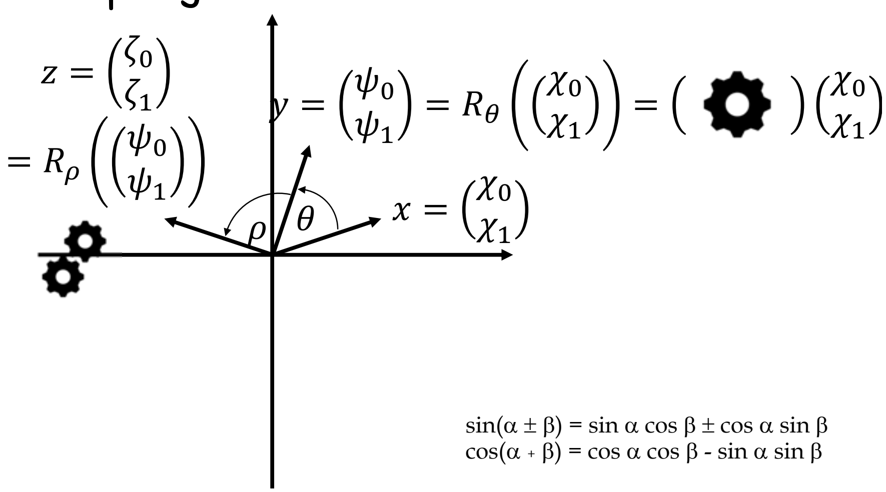
Opening Remarks

- Composing Rotations



Opening Remarks

- Composing Rotations



sin(α ± β) = sin α cos β ± cos α sin β
cos(α + β) = cos α cos β - sin α sin β

Observations

- Partitioned Matrix-Matrix Multiplication
- Properties Transposing a Product of Matrices
- Matrix-Matrix Multiplication with Special Matrices

Observations

- **Partitioned Matrix-Matrix Multiplication**

$$C = AB$$

- $$\begin{pmatrix} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \boxed{\gamma_{1,2}} & \gamma_{1,3} \\ \gamma_{2,0} & \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,0} & \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{pmatrix} =$$
$$\begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix}$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

$$(C_0 \mid C_1) = A(B_0 \mid B_1)$$

- $$\left(\begin{array}{cc|cc} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \boxed{\gamma_{1,2}} & \gamma_{1,3} \\ \gamma_{2,0} & \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,0} & \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{array}\right) = \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \left(\begin{array}{cc|cc} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{array}\right) =$$
$$\left(\begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} \\ \beta_{1,0} & \beta_{1,1} \\ \beta_{2,0} & \beta_{2,1} \\ \beta_{3,0} & \beta_{3,1} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{0,2} & \beta_{0,3} \\ \beta_{1,2} & \beta_{1,3} \\ \beta_{2,2} & \beta_{2,2} \\ \beta_{3,2} & \beta_{3,3} \end{pmatrix} \right)$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

$$\begin{pmatrix} C_0 \\ \hline C_1 \end{pmatrix} = \begin{pmatrix} A_0 \\ \hline A_1 \end{pmatrix} B$$

- $$\left(\begin{array}{cccc} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \boxed{\gamma_{1,2}} & \gamma_{1,3} \\ \hline \gamma_{2,0} & \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,0} & \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{array}\right) = \left(\begin{array}{cccc} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \hline \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{array}\right) \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix} =$$
$$\left(\begin{array}{c} \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix} \\ \hline \begin{pmatrix} \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix} \end{array}\right)$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

$$C = (A_0 \mid A_1) \begin{pmatrix} B_0 \\ \hline B_1 \end{pmatrix} = A_0 B_0 + A_1 B_1$$

- $$\begin{pmatrix} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \boxed{\gamma_{1,2}} & \gamma_{1,3} \\ \gamma_{2,0} & \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,0} & \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{pmatrix} = \left(\begin{array}{cc|cc} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{array}\right) \left(\begin{array}{cccc} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \hline \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{array}\right) =$$
$$\begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} \\ \alpha_{2,0} & \alpha_{2,1} \\ \alpha_{3,0} & \alpha_{3,1} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \end{pmatrix} + \begin{pmatrix} \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix}$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

$$\begin{pmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{pmatrix} = \begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix} \begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} = \begin{pmatrix} A_{00}B_{00} + A_{01}B_{10} & A_{01}B_{01} + A_{01}B_{11} \\ A_{10}B_{00} + A_{11}B_{10} & A_{10}B_{01} + A_{11}B_{11} \end{pmatrix}$$

- $$\begin{pmatrix} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \boxed{\gamma_{1,2}} & \gamma_{1,3} \\ \gamma_{2,0} & \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,0} & \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{pmatrix} = \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,0} & \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} & \beta_{0,3} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,0} & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix} =$$

$$\begin{pmatrix} \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} \\ \beta_{1,0} & \beta_{1,1} \end{pmatrix} + \begin{pmatrix} \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,2} & \alpha_{1,3} \end{pmatrix} \begin{pmatrix} \beta_{2,0} & \beta_{2,1} \\ \beta_{3,0} & \beta_{3,1} \end{pmatrix} & \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} \end{pmatrix} \begin{pmatrix} \beta_{0,2} & \beta_{0,3} \\ \beta_{1,2} & \beta_{1,3} \end{pmatrix} + \begin{pmatrix} \alpha_{0,2} & \alpha_{0,3} \\ \alpha_{1,2} & \alpha_{1,3} \end{pmatrix} \begin{pmatrix} \beta_{2,2} & \beta_{2,3} \\ \beta_{3,2} & \beta_{3,3} \end{pmatrix} \\ \begin{pmatrix} \alpha_{2,0} & \alpha_{2,1} \\ \alpha_{3,0} & \alpha_{3,1} \end{pmatrix} \begin{pmatrix} \beta_{0,0} & \beta_{0,1} \\ \beta_{1,0} & \beta_{1,1} \end{pmatrix} + \begin{pmatrix} \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{2,0} & \beta_{2,1} \\ \beta_{3,0} & \beta_{3,1} \end{pmatrix} & \begin{pmatrix} \alpha_{2,0} & \alpha_{2,1} \\ \alpha_{3,0} & \alpha_{3,1} \end{pmatrix} \begin{pmatrix} \beta_{0,2} & \beta_{0,3} \\ \beta_{1,2} & \beta_{1,3} \end{pmatrix} + \begin{pmatrix} \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \begin{pmatrix} \beta_{2,2} & \beta_{2,3} \\ \beta_{3,2} & \beta_{3,3} \end{pmatrix} \end{pmatrix}$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

Theorem 5.1 *Let $C \in \mathbb{R}^{m \times n}$, $A \in \mathbb{R}^{m \times k}$, and $B \in \mathbb{R}^{k \times n}$. Let*

- $m = m_0 + m_1 + \cdots m_{M-1}$, $m_i \geq 0$ *for* $i = 0, \ldots, M-1$;
- $n = n_0 + n_1 + \cdots n_{N-1}$, $n_j \geq 0$ *for* $j = 0, \ldots, N-1$; *and*
- $k = k_0 + k_1 + \cdots k_{K-1}$, $k_p \geq 0$ *for* $p = 0, \ldots, K-1$.

Partition

$$C = \left(\begin{array}{c|c|c|c} C_{0,0} & C_{0,1} & \cdots & C_{0,N-1} \\ \hline C_{1,0} & C_{1,1} & \cdots & C_{1,N-1} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline C_{M-1,0} & C_{M-1,1} & \cdots & C_{M-1,N-1} \end{array}\right), A = \left(\begin{array}{c|c|c|c} A_{0,0} & A_{0,1} & \cdots & A_{0,K-1} \\ \hline A_{1,0} & A_{1,1} & \cdots & A_{1,K-1} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline A_{M-1,0} & A_{M-1,1} & \cdots & A_{M-1,K-1} \end{array}\right),$$

$$\text{and } B = \left(\begin{array}{c|c|c|c} B_{0,0} & B_{0,1} & \cdots & B_{0,N-1} \\ \hline B_{1,0} & B_{1,1} & \cdots & B_{1,N-1} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline B_{K-1,0} & B_{K-1,1} & \cdots & B_{K-1,N-1} \end{array}\right),$$

with $C_{i,j} \in \mathbb{R}^{m_i \times n_j}$, $A_{i,p} \in \mathbb{R}^{m_i \times k_p}$, and $B_{p,j} \in \mathbb{R}^{k_p \times n_j}$. Then $C_{i,j} = \sum_{p=0}^{K-1} A_{i,p} B_{p,j}$.

Observations

- **Partitioned Matrix-Matrix Multiplication**

Example 5.2 Consider

$$A = \left(\begin{array}{rr|rr} -1 & 2 & 4 & 1 \\ 1 & 0 & -1 & -2 \\ 2 & -1 & 3 & 1 \\ 1 & 2 & 3 & 4 \end{array}\right), B = \left(\begin{array}{rrr} -2 & 2 & -3 \\ 0 & 1 & -1 \\ \hline -2 & -1 & 0 \\ 4 & 0 & 1 \end{array}\right), \text{ and } AB = \begin{pmatrix} -2 & -4 & 2 \\ -8 & 3 & -5 \\ -6 & 0 & -4 \\ 8 & 1 & -1 \end{pmatrix}:$$

If

$$A_0 = \begin{pmatrix} -1 & 2 \\ 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}, A_1 = \begin{pmatrix} 4 & 1 \\ -1 & -2 \\ 3 & 1 \\ 3 & 4 \end{pmatrix}, B_0 \begin{pmatrix} -2 & 2 & -3 \\ 0 & 1 & -1 \end{pmatrix}, \text{ and } B_1 = \begin{pmatrix} -2 & -1 & 0 \\ 4 & 0 & 1 \end{pmatrix}.$$

Observations

- **Partitioned Matrix-Matrix Multiplication**

Then

$$AB = \begin{pmatrix} A_0 & A_1 \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = A_0 B_0 + A_1 B_1:$$

$$\underbrace{\left(\begin{array}{rr|rr} -1 & 2 & 4 & 1 \\ 1 & 0 & -1 & -2 \\ 2 & -1 & 3 & 1 \\ 1 & 2 & 3 & 4 \end{array}\right)}_{A} \underbrace{\left(\begin{array}{rrr} -2 & 2 & -3 \\ 0 & 1 & -1 \\ \hline -2 & -1 & 0 \\ 4 & 0 & 1 \end{array}\right)}_{B}$$

$$= \underbrace{\begin{pmatrix} -1 & 2 \\ 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}}_{A_0} \underbrace{\begin{pmatrix} -2 & 2 & -3 \\ 0 & 1 & -1 \end{pmatrix}}_{B_0} + \underbrace{\begin{pmatrix} 4 & 1 \\ -1 & -2 \\ 3 & 1 \\ 3 & 4 \end{pmatrix}}_{A_1} \underbrace{\begin{pmatrix} -2 & -1 & 0 \\ 4 & 0 & 1 \end{pmatrix}}_{B_1}$$

$$= \underbrace{\begin{pmatrix} 2 & 0 & 1 \\ -2 & 2 & -3 \\ -4 & 3 & -5 \\ -2 & 4 & -5 \end{pmatrix}}_{A_0 B_0} + \underbrace{\begin{pmatrix} -4 & -4 & 1 \\ -6 & 1 & -2 \\ -2 & -3 & 1 \\ 10 & -3 & 4 \end{pmatrix}}_{A_1 B_1} = \underbrace{\begin{pmatrix} -2 & -4 & 2 \\ -8 & 3 & -5 \\ -6 & 0 & -4 \\ 8 & 1 & -1 \end{pmatrix}}_{AB}.$$

Observations

- Properties Transposing a Product of Matrices

Homework 5.2.3.1 Let $A = \begin{pmatrix} 2 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$. Compute

- $A^T A =$
- $A A^T =$
- $(AB)^T =$
- $A^T B^T =$
- $B^T A^T =$

Homework 5.2.3.2 Let $A \in \mathbb{R}^{m \times k}$ and $B \in \mathbb{R}^{k \times n}$. $(AB)^T = B^T A^T$.

Always/Sometimes/Never

Homework 5.2.3.3 Let A, B, and C be conformal matrices so that ABC is well-defined. Then $(ABC)^T = C^T B^T A^T$.

Always/Sometimes/Never

Observations

- Matrix-Matrix Multiplication with Special Matrices

Homework 5.2.4.9 Compute the following, using what you know about partitioned matrix-matrix multiplication:

$$\left(\begin{array}{cc|c} 1 & -1 & -2 \\ 0 & 2 & 3 \\ \hline 0 & 0 & 1 \end{array}\right)\left(\begin{array}{cc|c} -2 & 1 & -1 \\ 0 & 1 & 2 \\ \hline 0 & 0 & 1 \end{array}\right) =$$

Homework 5.2.4.14 Evaluate

- $\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}\begin{pmatrix} -1 & 1 & 2 \end{pmatrix} =$
- $\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\begin{pmatrix} 2 & 0 & -1 \end{pmatrix} =$
- $\left(\begin{array}{c|c} -1 & 2 \\ 1 & 0 \\ 2 & -1 \end{array}\right)\left(\begin{array}{ccc} -1 & 1 & 2 \\ \hline 2 & 0 & -1 \end{array}\right) =$
- $\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}\begin{pmatrix} 1 & -2 & 2 \end{pmatrix} =$
- $\left(\begin{array}{cc|c} -1 & 2 & 1 \\ 1 & 0 & -2 \\ 2 & -1 & 2 \end{array}\right)\left(\begin{array}{ccc} -1 & 1 & 2 \\ 2 & 0 & -1 \\ \hline 1 & -2 & 2 \end{array}\right) =$

Algorithms for Computing Matrix-Matrix Multiplication

- Lots of Loops
- Matrix-Matrix Multiplication by Columns
- Matrix-Matrix Multiplication by Rows
- Matrix-Matrix Multiplication with Rank-1 Updates

Algorithms for Computing Matrix-Matrix Multiplication

- Lots of Loops

$$C = AB \qquad \gamma_{i,j} = \sum_{p=0}^{k-1} \alpha_{i,p}\beta_{p,j}$$

- Consider the MATLAB function

```
for i = 1 : r_A
    for j = 1 : c_B
        for p = 1 : c_A
            C(i, j) = A(i ,p) * B(p ,j) + C(i ,j);
        end
    end
end
```

Algorithms for Computing Matrix-Matrix Multiplication

- **Matrix-Matrix Multiplication by Columns**

In Theorem 5.1 let us partition C and B by columns and not partition A. In other words, let $M = 1$, $m_0 = m$; $N = n$, $n_j = 1$, $j = 0, \ldots, n-1$; and $K = 1$, $k_0 = k$. Then

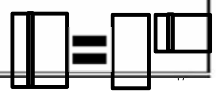
$$C = \left(c_0 \mid c_1 \mid \cdots \mid c_{n-1} \right) \quad \text{and} \quad B = \left(b_0 \mid b_1 \mid \cdots \mid b_{n-1} \right)$$

so that

$$\left(c_0 \mid c_1 \mid \cdots \mid c_{n-1} \right) = C = AB = A \left(b_0 \mid b_1 \mid \cdots \mid b_{n-1} \right) = \left(Ab_0 \mid Ab_1 \mid \cdots \mid Ab_{n-1} \right).$$

Homework 5.3.2.3

- $\begin{pmatrix} 1 & -2 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \left(\begin{array}{r|l} -1 & \\ 2 & \\ 1 & \end{array} \right) =$
- $\begin{pmatrix} 1 & -2 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \left(\begin{array}{r|r|l} -1 & 0 & \\ 2 & 1 & \\ 1 & -1 & \end{array} \right) =$
- $\begin{pmatrix} 1 & -2 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \left(\begin{array}{rr|r} -1 & 0 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & 2 \end{array} \right) =$



Algorithms for Computing Matrix-Matrix Multiplication

- **Matrix-Matrix Multiplication by Columns**

By moving the loop indexed by j to the outside in the algorithm for computing $C = AB + C$ we observe that

```
for j = 0,...,n-1                                          for j = 0,...,n-1
  for i = 0,...,m-1                                          for p = 0,...,k-1
    for p = 0,...,k-1                                          for i = 0,...,m-1
      γ_{i,j} := α_{i,p} β_{p,j} + γ_{i,j}  } c_j := Ab_j + c_j   or     γ_{i,j} := α_{i,p} β_{p,j} + γ_{i,j}  } c_j := Ab_j + c_j
    endfor                                                     endfor
  endfor                                                     endfor
endfor                                                     endfor
```

Exchanging the order of the two inner-most loops merely means we are using a different algorithm (dot product vs. AXPY) for the matrix-vector multiplication $c_j := Ab_j + c_j$.

Algorithms for Computing Matrix-Matrix Multiplication

- **Matrix-Matrix Multiplication by Columns**

Algorithm: $C := \text{GEMM\_UNB\_VAR1}(A, B, C)$

Partition $B \rightarrow \left(B_L \mid B_R \right)$, $C \rightarrow \left(C_L \mid C_R \right)$
where B_L has 0 columns, C_L has 0 columns

while $n(B_L) < n(B)$ **do**

Repartition

$\left(B_L \mid B_R \right) \rightarrow \left(B_0 \mid b_1 \mid B_2 \right)$, $\left(C_L \mid C_R \right) \rightarrow \left(C_0 \mid c_1 \mid C_2 \right)$
where b_1 has 1 column, c_1 has 1 column

$c_1 := Ab_1 + c_1$

Continue with

$\left(B_L \mid B_R \right) \leftarrow \left(B_0 \mid b_1 \mid B_2 \right)$, $\left(C_L \mid C_R \right) \leftarrow \left(C_0 \mid c_1 \mid C_2 \right)$

endwhile

Algorithms for Computing Matrix-Matrix Multiplication

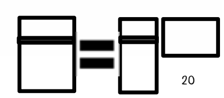
- **Matrix-Matrix Multiplication by Rows**

In Theorem 5.1 partition C and A by rows and do not partition B. In other words, let $M = m$, $m_i = 1$, $i = 0, \ldots, m-1$; $N = 1$, $n_0 = n$; and $K = 1$, $k_0 = k$. Then

$$C = \begin{pmatrix} \tilde{c}_0^T \\ \hline \tilde{c}_1^T \\ \hline \vdots \\ \hline \tilde{c}_{m-1}^T \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} \tilde{a}_0^T \\ \hline \tilde{a}_1^T \\ \hline \vdots \\ \hline \tilde{a}_{m-1}^T \end{pmatrix}$$

so that

$$\begin{pmatrix} \tilde{c}_0^T \\ \hline \tilde{c}_1^T \\ \hline \vdots \\ \hline \tilde{c}_{m-1}^T \end{pmatrix} = C = AB = \begin{pmatrix} \tilde{a}_0^T \\ \hline \tilde{a}_1^T \\ \hline \vdots \\ \hline \tilde{a}_{m-1}^T \end{pmatrix} B = \begin{pmatrix} \tilde{a}_0^T B \\ \hline \tilde{a}_1^T B \\ \hline \vdots \\ \hline \tilde{a}_{m-1}^T B \end{pmatrix}.$$



Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication by Rows

Homework 5.3.3.2

- $\left(\begin{array}{rrr} 1 & -2 & 2 \\ \hline & & \\ & & \end{array}\right)\left(\begin{array}{rrr} -1 & 0 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & 2 \end{array}\right) =$
- $\left(\begin{array}{rrr} 1 & -2 & 2 \\ \hline -1 & 2 & 1 \\ \hline & & \end{array}\right)\left(\begin{array}{rrr} -1 & 0 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & 2 \end{array}\right) =$
- $\left(\begin{array}{rrr} 1 & -2 & 2 \\ -1 & 2 & 1 \\ \hline 0 & 1 & 2 \end{array}\right)\left(\begin{array}{rrr} -1 & 0 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & 2 \end{array}\right) =$

Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication by Rows

In the algorithm for computing $C = AB + C$ the loop indexed by i can be moved to the outside so that

$$\begin{array}{l} \textbf{for } i = 0,\ldots,m-1 \\ \left.\begin{array}{l} \quad \textbf{for } j = 0,\ldots,n-1 \\ \quad\quad \textbf{for } p = 0,\ldots,k-1 \\ \quad\quad\quad \gamma_{i,j} := \alpha_{i,p}\beta_{p,j} + \gamma_{i,j} \\ \quad\quad \textbf{endfor} \\ \quad \textbf{endfor} \end{array}\right\} \tilde{c}_i^T := \tilde{a}_i^T B + \tilde{c}_i^T \\ \textbf{endfor} \end{array} \quad \text{or} \quad \begin{array}{l} \textbf{for } i = 0,\ldots,m-1 \\ \left.\begin{array}{l} \quad \textbf{for } p = 0,\ldots,k-1 \\ \quad\quad \textbf{for } j = 0,\ldots,n-1 \\ \quad\quad\quad \gamma_{i,j} := \alpha_{i,p}\beta_{p,j} + \gamma_{i,j} \\ \quad\quad \textbf{endfor} \\ \quad \textbf{endfor} \end{array}\right\} \tilde{c}_i^T := \tilde{a}_i^T B + \tilde{c}_i^T \\ \textbf{endfor} \end{array}$$

Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication by Rows

Algorithm: $C := \text{GEMM\_UNB\_VAR2}(A, B, C)$

Partition $A \rightarrow \left(\begin{array}{c} A_T \\ \hline A_B \end{array}\right)$, $C \rightarrow \left(\begin{array}{c} C_T \\ \hline C_B \end{array}\right)$

where A_T has 0 rows, C_T has 0 rows

while $m(A_T) < m(A)$ **do**

Repartition

$$\left(\begin{array}{c} A_T \\ \hline A_B \end{array}\right) \rightarrow \left(\begin{array}{c} A_0 \\ \hline a_1^T \\ \hline A_2 \end{array}\right), \left(\begin{array}{c} C_T \\ \hline C_B \end{array}\right) \rightarrow \left(\begin{array}{c} C_0 \\ \hline c_1^T \\ \hline C_2 \end{array}\right)$$

where a_1 has 1 row, c_1 has 1 row

$c_1^T := a_1^T B + c_1^T$

Continue with

$$\left(\begin{array}{c} A_T \\ \hline A_B \end{array}\right) \leftarrow \left(\begin{array}{c} A_0 \\ \hline a_1^T \\ \hline A_2 \end{array}\right), \left(\begin{array}{c} C_T \\ \hline C_B \end{array}\right) \leftarrow \left(\begin{array}{c} C_0 \\ \hline c_1^T \\ \hline C_2 \end{array}\right)$$

endwhile

Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication with Rank-1 Updates

In Theorem 5.1 partition A and B by columns and rows, respectively, and do not partition C. In other words, let $M = 1$, $m_0 = m$; $N = 1$, $n_0 = n$; and $K = k$, $k_p = 1$, $p = 0,\ldots,k-1$. Then

$$A = \left(a_0 \,\middle|\, a_1 \,\middle|\, \cdots \,\middle|\, a_{k-1} \right) \quad \text{and} \quad B = \left(\begin{array}{c} \tilde{b}_0^T \\ \hline \tilde{b}_1^T \\ \hline \vdots \\ \hline \tilde{b}_{k-1}^T \end{array}\right)$$

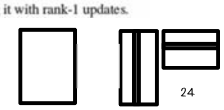
so that

$$C = AB = \left(a_0 \,\middle|\, a_1 \,\middle|\, \cdots \,\middle|\, a_{k-1} \right) \left(\begin{array}{c} \tilde{b}_0^T \\ \hline \tilde{b}_1^T \\ \hline \vdots \\ \hline \tilde{b}_{k-1}^T \end{array}\right) = a_0\tilde{b}_0^T + a_1\tilde{b}_1^T + \cdots + a_{k-1}\tilde{b}_{k-1}^T.$$

Notice that each term $a_p\tilde{b}_p^T$ is an outer product of a_p and $\tilde{b}_p$. Thus, if we start with $C := 0$, the zero matrix, then we can compute $C := AB + C$ as

$$C := a_{k-1}\tilde{b}_{k-1}^T + (\cdots + (a_p\tilde{b}_p^T + (\cdots + (a_1\tilde{b}_1^T + (a_0\tilde{b}_0^T + C))\cdots))\cdots),$$

which illustrates that $C := AB$ can be computed by first setting C to zero, and then repeatedly updating it with rank-1 updates.



Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication with Rank-1 Updates

Homework 5.3.4.1

- $\left(\begin{array}{c|cc} 1 & & \\ -1 & & \\ 0 & & \end{array}\right)\left(\begin{array}{ccc} -1 & 0 & 1 \\ \hline & & \\ \hline & & \end{array}\right) =$
- $\left(\begin{array}{c|c|c} & -2 & \\ & 2 & \\ & 1 & \end{array}\right)\left(\begin{array}{ccc} & & \\ \hline 2 & 1 & -1 \\ \hline & & \end{array}\right) =$
- $\left(\begin{array}{cc|c} & & 2 \\ & & 1 \\ & & 2 \end{array}\right)\left(\begin{array}{ccc} & & \\ \hline & & \\ \hline 1 & -1 & 2 \end{array}\right) =$
- $\begin{pmatrix} 1 & -2 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} -1 & 0 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & 2 \end{pmatrix} =$

Algorithms for Computing Matrix-Matrix Multiplication

- Matrix-Matrix Multiplication with Rank-1 Updates

Spark

Algorithm: $C := \text{GEMM\_UNB\_VAR3}(A, B, C)$

Partition $A \rightarrow \left(A_L \,\middle|\, A_R \right)$, $B \rightarrow \left(\begin{array}{c} B_T \\ \hline B_B \end{array}\right)$

where A_L has 0 columns, B_T has 0 rows

while $n(A_L) < n(A)$ **do**

Repartition

$$\left(A_L \,\middle|\, A_R \right) \rightarrow \left(A_0 \,\middle|\, a_1 \,\middle|\, A_2 \right), \left(\begin{array}{c} B_T \\ \hline B_B \end{array}\right) \rightarrow \left(\begin{array}{c} B_0 \\ \hline b_1^T \\ \hline B_2 \end{array}\right)$$

where a_1 has 1 column, b_1 has 1 row

$C := a_1 b_1^T + C$

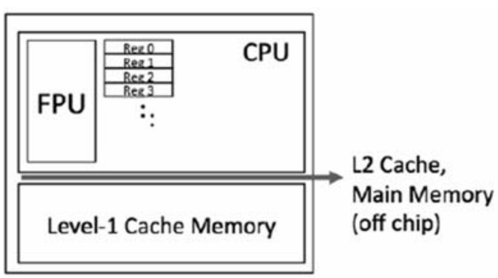
Continue with

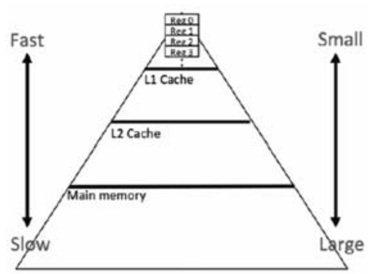
$$\left(A_L \,\middle|\, A_R \right) \leftarrow \left(A_0 \,\middle|\, a_1 \,\middle|\, A_2 \right), \left(\begin{array}{c} B_T \\ \hline B_B \end{array}\right) \leftarrow \left(\begin{array}{c} B_0 \\ \hline b_1^T \\ \hline B_2 \end{array}\right)$$

endwhile

Enrichment

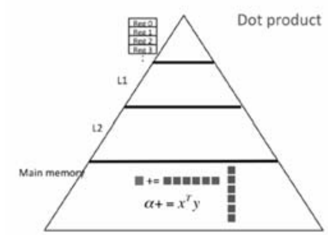
- Slicing and Dicing for Performance
  - CPU : Central Processing Unit
    - FPU : Floating Point Unit
    - Registers



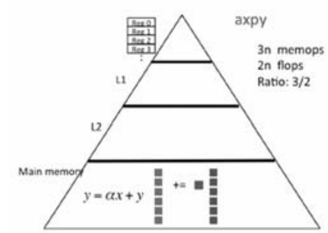


Enrichment

- Slicing and Dicing for Performance
  - Vector-Vector Computations

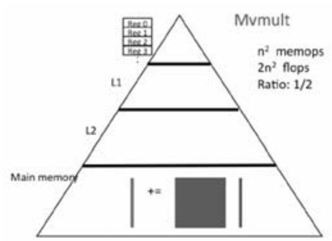


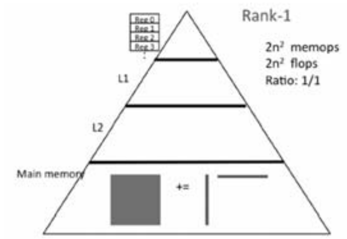
  - Matrix-Vector Computations



Enrichment

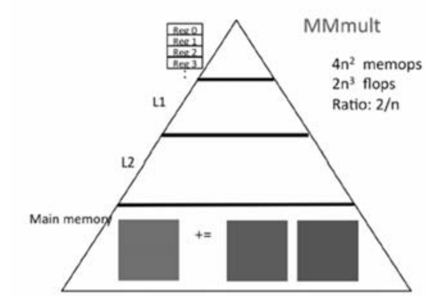
- Slicing and Dicing for Performance
  - Matrix-Vector Computations

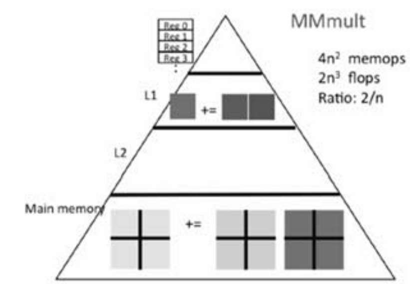




Enrichment

- Slicing and Dicing for Performance
  - Matrix-Matrix Computations





Questions and Answers