Matrix Products — (1/27)
1. Student Learning Objectives
   - SLOs: Matrix Products

2. Matrix Products
   - Motivation
   - Multiplication Mechanics

3. Suggested Problems
   - Suggested Problems 2.3
   - Lecture–Book Roadmap

4. Supplemental Material
   - Metacognitive Reflection
   - Problem Statements 2.3
After this lecture you should:

- Understand the Computational, and Linear Transformation Points-Of-View of Matrix Products
- Know that Matrix Multiplication is *Non-Commutative*
- Know that it is *not* always possible to multiply two matrices

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩
It is possible to express the numerical computation of (approximate) derivatives of a sampled function as a matrix-vector product $D\vec{u}$ where $\vec{u}$ is the function computed (sampled) at some number of points:

**Figure:** Sampled Function [Top], and Numerical Derivative [Bottom] for $n = 16$ [Left], and $n = 64$ [Right] sample points.
Those who have suffered through calculus wonder, **“What is this magic matrix which computes derivatives?!?”**

Let’s postpone the details (we need Taylor’s Theorem) of how to build such a matrix until a “bit” later... However, it has a very particular structure; with lots of zeros:

**Figure:** The structure of the “differentiation matrix.” It turns out that the approximation error in the computations is proportional to the square of the distance between the points. That means if we double the number of points (cut the distance in half), we reduce the error by a factor of $\frac{1}{4}$. 
We can get higher quality approximations by either adding more points; or putting more work into crafting the approximation matrix:

**Figure:** [Left], Numerical Derivative for \( n = 1024 \) points; and [Right] The structure of a “differentiation matrix” which produces errors proportional to the distance between the points to the power 4. That means if we double the number of points, we reduce the error by a factor of \( \frac{1}{16} \).
OK, say you have invested all that effort into building these differentiation matrices... and now some evil professor person comes along and wants second derivatives.

\[ D D \vec{u} \] will do the trick!

**Figure:** Sampled Function [Top], and Numerical 2nd Derivative [Bottom] for \( n = 64 \) [Left], and \( n = 256 \) [Right] sample points.
We can describe a sequence of linear transformations e.g. the Scaling ($M_s$) of an Orthogonally Projected ($M_o$) Reflection ($M_r$) of a Horizontally Sheared ($M_{hs}$) geometric object as a sequence of matrix-vector multiplications:

$$M_s M_o M_r M_{hs} \vec{u}$$

In signal analysis (applications JPEG, MPEG compression and beyond) we can express the discrete cos-transform ($DCT$) (and its inverse) as matrix multiplications; and (certain linear) filters can also be expressed as matrix multiplications; so it is reasonable to compute things like

$$M_{\cos^{-1}} M_{\text{filter}} M_{\cos} \vec{u}$$
Matrix Multiplication :: Computational P.O.V.

Let $B \in \mathbb{R}^{n \times p}$, and $A \in \mathbb{R}^{q \times m}$:

- The product $BA$ is defined if and only if $p = q$; when it is defined $C = BA$ gives a matrix $C \in \mathbb{R}^{n \times m}$. The entry in row $i$, column $j$ of $C$ is given by
  
  $$c_{ij} = \sum_{k=1}^{p} b_{ik} a_{kj}.$$  
  Dot product of $i$th row of $B$, and $j$th column of $A$

- The product $AB$ is defined if and only if $m = n$; when it is defined $D = AB$ gives a matrix $D \in \mathbb{R}^{q \times p}$. The entry in row $i$, column $j$ of $D$ is given by
  
  $$d_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$  
  Dot product of $i$th row of $A$, and $j$th column of $B$
Let $B \in \mathbb{R}^{n \times p}$, and $A \in \mathbb{R}^{q \times m}$; then the product $BA$ is defined as the matrix of the linear transformation $T(\vec{x}) = B(A\vec{x})$. This means that $T(\vec{x}) = B(A\vec{x}) = (BA)\vec{x}$, $\forall \vec{x} \in \mathbb{R}^m$; the product $BA \in \mathbb{R}^{n \times m}$. 
The Columns of the Matrix Product

Let $B$ be an $n \times p$ matrix, and $A$ a $p \times m$ matrix with columns $\vec{a}_1$, $\vec{a}_2$, $\ldots$, $\vec{a}_m \in \mathbb{R}^p$, then

$$BA = B \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \ldots & \vec{a}_m \end{bmatrix} = \begin{bmatrix} B\vec{a}_1 & B\vec{a}_2 & \ldots & B\vec{a}_m \end{bmatrix}.$$ 

To find $BA$, we multiply the columns of $A$ by $B$, and collect the resulting vectors as columns in the resulting matrix.
Matrix Multiplication is Non-Commutative

In general $BA \neq AB$.

In the rare cases when $AB = BA$; the we say that the matrices commute.

Example: Let $(A \in \mathbb{R}^{3 \times 2}, B \in \mathbb{R}^{2 \times 3} \Rightarrow AB \in \mathbb{R}^{3 \times 3}, BA \in \mathbb{R}^{2 \times 2})$

\[
A = \begin{bmatrix}
5 & -4 \\
0 & -1 \\
3 & 5 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
3 & 2 & 4 \\
5 & -5 & 5 \\
\end{bmatrix};
\]

then

\[
AB = \begin{bmatrix}
-5 & 30 & 0 \\
-5 & 5 & -5 \\
34 & -19 & 37 \\
\end{bmatrix}, \quad \text{and} \quad BA = \begin{bmatrix}
27 & 6 \\
40 & 10 \\
\end{bmatrix}.
\]
Another Demonstration of the Non-Commutative Property

**Example:** Let \((A, B \in \mathbb{R}^{3 \times 3} \Rightarrow AB, BA \in \mathbb{R}^{3 \times 3})\)

\[
A = \begin{bmatrix}
2 & -1 & 2 \\
3 & 2 & -5 \\
3 & -4 & -2
\end{bmatrix}, \quad B = \begin{bmatrix}
-5 & 2 & -5 \\
-4 & -2 & -1 \\
4 & 5 & -1
\end{bmatrix};
\]

then

\[
AB = \begin{bmatrix}
2 & 16 & -11 \\
-43 & -23 & -12 \\
-7 & 4 & -9
\end{bmatrix}, \quad \text{and} \quad BA = \begin{bmatrix}
-19 & 29 & -10 \\
-17 & 4 & 4 \\
20 & 10 & -15
\end{bmatrix}.
\]
Multiplying by the Identity Matrix

If $A \in \mathbb{R}^{m \times n}$, then

$$I_m A = A, \quad \text{and} \quad A I_n = A$$

where $I_m$ is the $m \times m$ identity matrix, and $I_n$ the $n \times n$ identity matrix.
Let $A \in \mathbb{R}^{n \times p}$, $B \in \mathbb{R}^{p \times q}$, and $C \in \mathbb{R}^{q \times m}$; then clearly

- The products $A B \in \mathbb{R}^{n \times q}$ and $B C \in \mathbb{R}^{p \times m}$ make sense.
- Given the resulting sizes, we can take the results and compute $(AB)C \in \mathbb{R}^{n \times m}$, and $A(BC) \in \mathbb{R}^{n \times m}$.
- So, yeah, they are the same sizes... but $A(BC) \overset{??}{=} (AB)C$

Indeed, they are... and the Linear Transformation P.O.V. of the matrix product helps: — we have

$$T_1(\vec{x}) = ((AB)C)\vec{x}, \quad \text{and} \quad T_2(\vec{x}) = (A(BC))\vec{x}$$
Matrix Multiplication is Associative

... and using the Linear Transformation P.O.V. of the matrix product gives:

\[ T_1(\vec{x}) = ((AB)C)\vec{x} = (AB)(C\vec{x}) = A(B(C\vec{x})) \]

and

\[ T_2(\vec{x}) = (A(BC))\vec{x} = A((BC)\vec{x}) = A(B(C\vec{x})) \]

If that makes you unhappy, you can use the computational P.O.V.
Matrix Multiplication is Associative

Let \( A \in \mathbb{R}^{n \times p}, \ B \in \mathbb{R}^{p \times q}, \) and \( C \in \mathbb{R}^{q \times m}, \) then

\[
(AB)_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \quad (BC)_{k\ell} = \sum_{j=1}^{q} b_{kj} c_{j\ell}
\]

\[
((AB)C)_{i\ell} = \sum_{j=1}^{q} (AB)_{ij} c_{j\ell} = \sum_{j=1}^{q} \left[ \sum_{k=1}^{p} a_{ik} b_{kj} \right] c_{j\ell} = \sum_{j=1}^{q} \sum_{k=1}^{p} a_{ik} b_{kj} c_{j\ell}
\]

\[
A(BC)_{i\ell} = \sum_{k=1}^{p} a_{ik} (BC)_{k\ell} = \sum_{k=1}^{p} a_{ik} \left[ \sum_{j=1}^{q} b_{kj} c_{j\ell} \right] = \sum_{k=1}^{p} \sum_{j=1}^{q} a_{ik} b_{kj} c_{j\ell}
\]

... and since order of summation does not matter, they are equal.

Now we’re all smiles(?!)

Peter Blomgren, (blomgren.peter@gmail.com)
Distributive Property for Matrices

If $A, B \in \mathbb{R}^{n \times p}$ and $C, D \in \mathbb{R}^{p \times m}$, then

$$A(C + D) = AC + AD,$$

$$\text{and}$$

$$(A + B)C = AC + BC.$$ 

This can be shown either using the Linear Transform, or the Computational P.O.V. (have “fun!”)
Scaling

If $A \in \mathbb{R}^{n \times p}$, $B \in \mathbb{R}^{p \times m}$, $k \in \mathbb{R}$, then

$$(kA)B = A(kB) = k(AB)$$
Available on Learning Glass videos:
2.3 — 1, 3, 5, 7, 13, 17, 19, 27, 28, 33, 37
### Lecture – Book Roadmap

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Book, [GS5–]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>§2.2</td>
</tr>
<tr>
<td>1.2</td>
<td>§1.1, §1.3, §2.1, §2.3</td>
</tr>
<tr>
<td>1.3</td>
<td>§1.1, §1.2, §1.3, §2.1, §2.3</td>
</tr>
<tr>
<td>1.4</td>
<td>§1.1–§1.3, §2.1–§2.3</td>
</tr>
<tr>
<td>2.1</td>
<td>§8.1, §8.2*, §2.5*</td>
</tr>
<tr>
<td>2.2</td>
<td>§8.1, §8.2*, §4.2*, §4.4*</td>
</tr>
<tr>
<td>2.3</td>
<td>§2.4</td>
</tr>
</tbody>
</table>

§2.5* (p.86–88) “Calculating $A^{-1}$ by Gauss-Jordan Elimination”

§4.2* (p.207) “Projection Onto a Line” – (p.210) end of “Example 2”

§4.4* Example 1, Example 3

§8.2* We will talk about “Basis” / “Bases” soon… don’t worry about those concepts… yet.
## Metacognitive Exercise — Thinking About Thinking & Learning

<table>
<thead>
<tr>
<th>I know / learned</th>
<th>Almost there</th>
<th>Huh?!?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right After Lecture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>After Thinking / Office Hours / SI-session</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>After Reviewing for Midterm/Final</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩

Matrix Products
(2.3.1) Compute (if possible) the matrix product \((i)\) column-by-column, and \((ii)\) entry-by-entry.

\[
\begin{bmatrix}
1 & 1 \\
0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix}
\]

(2.3.3) Compute (if possible) the matrix product \((i)\) column-by-column, and \((ii)\) entry-by-entry.

\[
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix}
\]
(2.3.5), (2.3.7)

(2.3.5) Compute (if possible) the matrix product (i) column-by-column, and (ii) entry-by-entry.

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix}
\]

(2.3.7) Compute (if possible) the matrix product (i) column-by-column, and (ii) entry-by-entry.

\[
\begin{bmatrix}
1 & 0 & -1 \\
0 & 1 & 1 \\
1 & -1 & -2
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 3 \\
3 & 2 & 1 \\
2 & 1 & 3
\end{bmatrix}
\]
(2.3.13) Compute (if possible) the matrix product (i) column-by-column, and (ii) entry-by-entry.

\[
\begin{bmatrix}
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & k
\end{bmatrix}
\begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\]

(2.3.17) Find all matrices that commute with

\[
A = \begin{bmatrix}
1 & 0 \\
0 & 2
\end{bmatrix}
\]
(2.3.19) Find all matrices that commute with

\[ A = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} \]

(2.3.27) Prove the *distributive laws* for matrices:

\[ A(C + D) = AC + AD, \quad \text{and} \quad (A + B)C = AC + BC. \]

(2.3.28) Consider an \( n \times p \) matrix \( A \), a \( p \times m \) matrix \( B \), and a scalar \( k \). Show that

\[ (kA)B = A(kB) = k(AB) \]
(2.3.33) For the given matrix \( A \), compute \( A^2 = AA \), \( A^3 = AAA \), and \( A^4 \). Describe the emerging pattern, and use it to find \( A^{1001} \).
— Interpret in terms of rotations, reflections, shears, and orthogonal projections.

\[
A = \begin{bmatrix}
-1 & 0 \\
0 & -1
\end{bmatrix}
\]

(2.3.37) For the given matrix \( A \), compute \( A^2 = AA \), \( A^3 = AAA \), and \( A^4 \). Describe the emerging pattern, and use it to find \( A^{1001} \).
— Interpret in terms of rotations, reflections, shears, and orthogonal projections.

\[
A = \begin{bmatrix}
1 & 0 \\
-1 & 1
\end{bmatrix}
\]