Session 0: Linear Algebra and Calculus Review

ENGR 510: Engineering Optimization

Dr. Daniel R. Herber

daniel.herber@colostate.edu

Colorado State University – Department of Systems Engineering

https://colostate.instructure.com/courses/187606
⚠ Disclaimer: This document is for educational, noncommercial purposes only. Reproduction, distribution, republication, and/or retransmission of material contained within this document are prohibited unless the prior written permission of Daniel R. Herber (daniel.herber@colostate.edu) has been obtained.

This document may contain copyrighted material whose use has not been specifically authorized by the copyright owner. The creator of this document is making them available in an effort to advance the understanding of the educational material contained within. We believe that this constitutes a “fair use” of the copyrighted material as provided for in section 107 of the US Copyright Law. If you wish to use this copyrighted material for purposes of your own that go beyond “fair use”, you must obtain permission from the copyright owner.

The views and opinions expressed in this document are those of the creator and do not necessarily reflect the official policy or position of any other entity.

While the information contained within this document is periodically updated, no guarantee is given that the information provided in this document is correct, complete, and/or up-to-date. If you find any errors in this document, please email daniel.herber@colostate.edu. Thanks!
Outline

1. Single Variable Calculus
2. Linear Algebra
3. Multivariable and Matrix Calculus
4. Summary
Intermediate and Extreme Value Theorems

Theorem (Intermediate Value)

If \( f \) is continuous on \([a, b]\) and \( k \) is any number between \( f(a) \) and \( f(b) \), then there is at least one \( c \) in the interval \((a, b)\) such that \( f(c) = k \).

- Consider \( f(x) = x^5 - 2x^3 - 2 = 0 \) and the interval \([0, 2]\), then \( f(0) = -2 \) and \( f(2) = 14 \), so \( f(c) = 0 \) for at least one \( c \) in \((0, 2)\).

Theorem (Extreme Value)

If \( f \) is continuous on a bounded closed interval \([a, b]\), then on that interval \( f \) takes on both a maximum value \( M \) and a minimum value \( m \).

- Consider \( f(x) = x^4 - 3x^3 - 1 \) on \([-2, 2]\), then the maximum value is \( M = 39 \) and minimum value is \( m = -9 \).
Rolle’s Theorem

Theorem (Rolle’s)

Suppose that $f$ is differentiable on the open interval $(a, b)$ and continuous on the closed interval $[a, b]$. If $f(a) = f(b)$, then there is at least one $c$ in $(a, b)$ for which:

$$f'(c) = 0$$

Consider $f(x) = x^2 + 1$ on $[-1, 1]$, then $f(-1) = f(1) = 2$, so by Rolle’s Theorem, we know that $f'(c) = 0$ for some $c \in (-1, 1)$, which is $c = 0$.
Mean Value Theorem

Theorem (Mean Value)

If $f$ is differentiable on the open interval $(a, b)$ and continuous on the closed interval $[a, b]$, then there is at least one $c$ in $(a, b)$ for which:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

(2)

- Consider $f(x) = x^2 + 1$ on $[1, 4]$, then knowing $f'(x) = 2x$:

$$\frac{f(b) - f(a)}{b - a} = \frac{17 - 2}{4 - 1} = 5 = f'(c) \implies c = 2.5$$

(3)

which lies in the interval $(1, 4)$
→ Derivative Definition

**Definition (Derivative)**

*Given a function $f$, the **derivative** of $f$ is the function $f'$ defined as:*

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$  \hspace{1cm} (4)

*provided the limit exists. The domain of $f'(x)$ is the set of all points where the defining limit exists, that is, all $x$ for which $f$ is differentiable.*
Product Rule

Definition (Product Rule)

For two functions $u$ and $v$, the product rule is, using Lagrange's notation:

$$(uv)' = u'v + uv'$$  \hspace{1cm} (5)

or in Leibniz’s notation is:

$$\frac{d}{dx}(uv) = \frac{du}{dx}v + u\frac{dv}{dx}$$  \hspace{1cm} (6)

- Consider $f(x) = (3x - 2x^2)(5 + 4x)$, the following the product rule:

$$f'(x) = \frac{d}{dx}(3x - 2x^2)(5 + 4x) + (3x - 2x^2)\frac{d}{dx}(5 + 4x)$$  \hspace{1cm} (7a)

$$= (3 - 4x)(5 + 4x) + (3x - 2x^2)(0 + 4)$$  \hspace{1cm} (7b)

$$= -24x^2 + 4x + 15$$  \hspace{1cm} (7c)


**→ Quotient Rule**

**Definition (Quotient Rule)**

Let \( f(x) = \frac{u}{v} \) where both \( u \) and \( v \) are differentiable and \( v(x) \neq 0 \). The quotient rule states that the derivative of \( f(x) \) is:

\[
    f' = \frac{u'v - uv'}{v^2}
\]  

(8)

• **Consider** \( f(x) = \frac{x^2 + 6}{2x - 7} \), then following the quotient rule:

\[
    f'(x) = \frac{\frac{d}{dx}(x^2 + 6)(2x - 7) - (x^2 + 6)\frac{d}{dx}(2x - 7)}{(2x - 7)^2}
\]  

(9a)

\[
    = \frac{(2x)(2x - 7) - (x^2 + 6)(2)}{(2x - 7)^2}
\]  

(9b)

\[
    = \frac{2(x^2 - 7x - 6)}{(2x - 7)^2}
\]  

(9c)
### Chain Rule

**Definition (Chain Rule)**

If $h = f \circ g$ is the function such that $h(x) = f(g(x))$ for every $x$, then the chain rule is, in Lagrange’s notation:

$$h'(x) = f'(g(x))g'(x) \quad (10)$$

or in Leibniz’s notation:

$$\frac{dh}{dx} = \frac{df}{dg} \frac{dg}{dx} \quad (11)$$

- Consider $h(x) = \sin(3x^2 + x)$, then following the chain rule with $g(x) = 3x^2 + x$:

  $$h'(x) = \frac{d\sin(g)}{dg} \frac{d(3x^2 + x)}{dx} \quad (12a)$$

  $$= \cos(g)(6x + 1) \quad (12b)$$

  $$= \cos(3x^2 + x)(6x + 1) \quad (12c)$$
L'Hôpital’s Rule

**Definition (L'Hôpital's Rule)**

For functions \( f \) and \( g \) which are differentiable on an open interval \( I \), except possibly at a point \( c \) in \( I \), if \( \lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0 \) or \( \pm \infty \) and \( g'(x) \neq 0 \) for all \( x \) in \( I \) with \( x \neq c \), and \( \lim_{x \to c} f'(x)/g'(x) \) exists, then:

\[
\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}
\]  

(13)

- 🔄 Consider the following example using L'Hôpital's rule:

\[
\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{\cos(x)}{1} = 1
\]  

(14)
First Fundamental Theorem of Calculus

**Theorem (First Fundamental Theorem of Calculus)**

Let $f$ be a continuous real-valued function defined on a closed interval $[a, b]$. Let $F$ be the function defined, for all $x$ in $[a, b]$, by:

$$F(x) = \int_a^x f(t)dt$$  \hspace{1cm} (15)

Then $F$ is uniformly continuous on $[a, b]$ and differentiable on the open interval $(a, b)$, and:

$$F'(x) = \frac{d}{dx} \int_a^x f(t)dt = f(x)$$  \hspace{1cm} (16)

for all $x$ in $(a, b)$.
Second Fundamental Theorem of Calculus

Theorem (Second Fundamental Theorem of Calculus)

Let $f$ be a real-valued function on a closed interval $[a, b]$ and $F$ an antiderivative of $f$ in $(a, b)$. If $f$ is Riemann integrable on $[a, b]$, then:

$$
\int_{a}^{b} f(x)dx = F(b) - F(a) \quad (17)
$$
### Table of Derivatives

<table>
<thead>
<tr>
<th>$y = f(x)$</th>
<th>$\frac{dy}{dx} = f'(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k, \text{ any constant}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$x$</td>
<td>$1$</td>
</tr>
<tr>
<td>$x^n, \text{ any constant } n$</td>
<td>$nx^{n-1}$</td>
</tr>
<tr>
<td>$e^x$</td>
<td>$e^x$</td>
</tr>
<tr>
<td>$e^{kx}$</td>
<td>$ke^{kx}$</td>
</tr>
<tr>
<td>$\ln(x) = \log_e(x)$</td>
<td>$\frac{1}{x}$</td>
</tr>
<tr>
<td>$\sin(x)$</td>
<td>$\cos(x)$</td>
</tr>
<tr>
<td>$\sin(kx)$</td>
<td>$k \cos(kx)$</td>
</tr>
<tr>
<td>$\cos(x)$</td>
<td>$-\sin(x)$</td>
</tr>
<tr>
<td>$\cos(kx)$</td>
<td>$-k \sin(kx)$</td>
</tr>
<tr>
<td>$\tan(x) = \frac{\sin(x)}{\cos(x)}$</td>
<td>$\sec^2(x)$</td>
</tr>
<tr>
<td>$\tan(kx)$</td>
<td>$k \sec^2(kx)$</td>
</tr>
<tr>
<td>$\csc(x) = \frac{1}{\sin(x)}$</td>
<td>$-\csc(x) \cot(x)$</td>
</tr>
<tr>
<td>$\sec(x) = \frac{1}{\cos(x)}$</td>
<td>$\sec(x) \tan(x)$</td>
</tr>
<tr>
<td>$\cot(x) = \frac{\cos(x)}{\sin(x)}$</td>
<td>$-\csc^2(x)$</td>
</tr>
<tr>
<td>$\sin^{-1}(x)$</td>
<td>$\frac{1}{\sqrt{1-x^2}}$</td>
</tr>
<tr>
<td>$\cos^{-1}(x)$</td>
<td>$-\frac{1}{\sqrt{1-x^2}}$</td>
</tr>
<tr>
<td>$\tan^{-1}(x)$</td>
<td>$\frac{1}{1+x^2}$</td>
</tr>
<tr>
<td>$\cosh(x)$</td>
<td>$\sinh(x)$</td>
</tr>
<tr>
<td>$\sinh(x)$</td>
<td>$\cosh(x)$</td>
</tr>
<tr>
<td>$\cosh^{-1}(x)$</td>
<td>$\frac{1}{\sqrt{x^2-1}}$</td>
</tr>
<tr>
<td>$\sinh^{-1}(x)$</td>
<td>$\frac{1}{\sqrt{x^2+1}}$</td>
</tr>
<tr>
<td>$\tanh^{-1}(x)$</td>
<td>$\frac{1}{1-x^2}$</td>
</tr>
</tbody>
</table>
### Table of Integrals

<table>
<thead>
<tr>
<th>$f(x)$</th>
<th>$\int f(x),dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$, any constant</td>
<td>$kx + c$</td>
</tr>
<tr>
<td>$x$</td>
<td>$\frac{x^2}{2} + c$</td>
</tr>
<tr>
<td>$x^n$, any $n \neq 1$</td>
<td>$\frac{x^{n+1}}{n+1} + c$</td>
</tr>
<tr>
<td>$x^{-1}$</td>
<td>$\ln</td>
</tr>
<tr>
<td>$e^x$</td>
<td>$e^x + c$</td>
</tr>
<tr>
<td>$e^{-x}$</td>
<td>$-e^{-x} + c$</td>
</tr>
<tr>
<td>$e^{kx}$</td>
<td>$\frac{e^{kx}}{k} + c$</td>
</tr>
<tr>
<td>$\sin(x)$</td>
<td>$-\cos(x) + c$</td>
</tr>
<tr>
<td>$\cos(x)$</td>
<td>$\sin(x) + c$</td>
</tr>
<tr>
<td>$\sin(kx)$</td>
<td>$-\frac{\cos(kx)}{k} + c$</td>
</tr>
<tr>
<td>$\cos(kx)$</td>
<td>$\frac{\sin(kx)}{k} + c$</td>
</tr>
<tr>
<td>$\tan(kx)$</td>
<td>$\frac{1}{k} \ln</td>
</tr>
</tbody>
</table>
References for Single Variable Calculus

- Some references include:
  - Active Calculus\(^1\) by Boelkins, Austin, and Schlicker (free)
  - Calculus of One Variable by Hirst (CSU available)

- Some tools include:
  - Matlab\(^2\)
  - Python SymPy\(^3\)
  - WolframAlpha\(^4\)

---

\(^1\) https://activecalculus.org/single/book-1.html — accompanying website
\(^2\) www.mathworks.com/help/symbolic/calculus.html
\(^3\) docs.sympy.org/latest/tutorial/calculus.html
\(^4\) www.wolframalpha.com/
Linear Algebra
Matrix Definition

Definition (Matrix)

An \( n \times m \) matrix is a collection of real or complex numbers \( a_{ij} \) organized in a rectangular array with \( n \) rows and \( m \) columns:

\[
A = \{a_{ij}\} = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1m} \\
a_{21} & a_{22} & \cdots & a_{2m} \\
& \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nm}
\end{bmatrix}
\]

(18)

Consider the following example of a matrix:

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

(19a)

\[
a_{11} = 1 \quad a_{13} = 3 \quad a_{23} = 4
\]

(19b)
Transposed Matrix

**Definition (Transposed Matrix)**

The transposed matrix is obtained by interchanging columns and rows:

$$[A^\top]_{ij} = a_{ji}$$

(20)

- Consider the following example transposing a matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix} \quad \mathbf{A}^\top = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \end{bmatrix}$$

(21)
→ Symmetric Matrix

Definition (Symmetric Matrix)

A symmetric matrix $A_S$ is equal to its transpose:

$$A_S = A_S^T$$  \hspace{1cm} (22)

Consider the following example of a symmetric matrix:

$$Q = Q^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 3 & 0 & 4 \end{bmatrix}$$  \hspace{1cm} (23)
Matrix Multiplication

Definition (Matrix Product)

If $A$ is an $n \times m$ matrix and $B$ is an $m \times p$ matrix, then the **matrix product** $AB = C$ is defined to be the $n \times p$ matrix:

$$[C]_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj} \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, p \quad (24)$$

- Consider the following example of matrix multiplication:

\[
\begin{pmatrix}
1 & 0 & 2 \\
3 & 1 & 0 \\
5 & -1 & 2
\end{pmatrix}
\begin{pmatrix}
2 & -1 \\
5 & 1 \\
-2 & 0
\end{pmatrix}
= \begin{pmatrix}
-2 & -1 \\
11 & -2 \\
1 & -6
\end{pmatrix}
\quad (25)
\]

with $c_{21}$ computed with:

\[
3 \times 2 + 1 \times 5 + 0 \times -2 = 11 \quad (26)
\]
Diagonal Matrices

**Definition (Diagonal Matrix)**

A diagonal matrix is a square matrix, where all off-diagonal elements are zero:

\[
\text{diag}(\alpha_1, \cdots, \alpha_n) = \begin{bmatrix}
\alpha_1 & 0 & \cdots & 0 \\
0 & \alpha_2 & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \alpha_n
\end{bmatrix}
\]  

(27)

- Consider the following example of a diagonal matrix:

\[
A = \text{diag}(-1, 2, 0) = \begin{bmatrix}
-1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]  

(28)
→ Block Diagonal Matrices

**Definition (Block Diagonal Matrix)**

A #block diagonal matrix has the following form:

\[
\text{diag}(A_1, \cdots, A_n) = \begin{bmatrix}
A_1 & 0 & \cdots & 0 \\
0 & A_2 & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & A_n
\end{bmatrix}
\]  

(29)

- Consider the following example of a block diagonal matrix:

\[
A = \text{diag}\left(\begin{bmatrix}1 & 2 \\ 3 & 4\end{bmatrix}, \begin{bmatrix}-3\end{bmatrix}\right) = \begin{bmatrix}1 & 2 & 0 \\
3 & 4 & 0 \\
0 & 0 & -3\end{bmatrix}
\]  

(30)
Trace

Definition (Trace of a Square Matrix)

The \textit{trace} of a square matrix (a matrix with equal number of rows and columns) is the sum of the diagonal elements:

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$$ (31)

- Consider the following example of the trace of a matrix:

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & -3 \end{bmatrix} \quad \text{tr}(A) = 1 + 4 + (-3) = 2$$ (32)
Determinant — Geometric Interpretation

**Definition (Determinant)**

Let $v_1, \cdots, v_n$ be vectors in $\mathbb{R}^n$, $P$ be the parallelepiped determined by these vectors, and $A$ be the matrix with rows $v_1, \cdots, v_n$. Then, the absolute value of the determinant of $A$ is the volume of $P$:

$$|\text{det}(A)| = \text{vol}(P)$$

(33)

- Example on Slide 24
- Some properties of the determinant:

$$\text{det}(A^{-1}) = \frac{1}{\text{det}(A)}$$

(34a)

$$\text{det}(A^\top) = \text{det}(A)$$

(34b)

$$\text{det}(cA) = c^n \text{det}(A) \text{ for } A \text{ of size } n \times n$$

(34c)

---

1 textbooks.math.gatech.edu/ila/determinants-volumes.html
Example of the Determinant

- For a $2 \times 2$ matrix, the determinant represents the area of the parallelogram defined by the vectors illustrated below:

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$
→ Rank

**Definition (Rank of a Matrix)**

The rank of a matrix is defined as the number of linearly independent columns or rows. If $\mathbf{A}$ is an $n \times m$ matrix, then $r$ is the rank if any $r + 1$ columns or rows of $\mathbf{A}$ are linearly dependent but any $r$ columns or rows are linearly independent. Denote $\text{rank}(\mathbf{A}) = r$.

- This definition means that if $[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r]$ are any $r$ columns or rows of $\mathbf{A}$, then the following implication is true:

  $$\text{if } \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \cdots + \alpha_n \mathbf{a}_r = \mathbf{0}, \text{ then } \alpha_i = 0 \text{ for all } i \tag{36}$$

- Consider the following example:

  $$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{bmatrix} \tag{37a}$$

  $$\text{rank}(\mathbf{A}) = 2 \neq 3 \text{ because } (1) \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + (1) \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} + (1) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0} \tag{37b}$$
Identity Matrix

**Definition (Identity Matrix)**

_The identity matrix is a square matrix of size \( n \times n \):_

\[
I = I_n = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}
\]  

(38)
Eigenvalues and Eigenvectors

Definition (Eigenvalues and Eigenvectors)

The characteristic polynomial for a square $n \times n$ matrix $A$ is defined as:

$$a(s) = \text{det}(sI_n - A) = s^n + a_{n-1}s^{n-1} + \cdots + a_1s + a_0$$ (39)

where the $n$ roots of the characteristic polynomial (i.e., $a(s) = 0$) are called the eigenvalues of $A$.

For each of the eigenvalues $\lambda_i$, there exists a nonzero vector $\mathbf{v}_i$, the eigenvector, such that the eigenvector equation is satisfied:

$$A\mathbf{v}_i = \lambda_i\mathbf{v}_i$$ (40)

• Example on Slide 28
Example of Eigenvalues and Eigenvectors

- Consider the following example:

\[ A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \tag{41a} \]

\[ a(s) = \det(sI_n - A) = \begin{bmatrix} s & -1 \\ 2 & 3 + s \end{bmatrix} = s^2 + 3s + 2 \tag{41b} \]

\[ a(s) = s^2 + 3s + 2 = (s + 1)(s + 2) = 0 \implies \lambda_1 = -1, \lambda_2 = -2 \tag{41c} \]

\[ v_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \text{ because } \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = -1 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \tag{41d} \]

\[ v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \text{ because } \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = -2 \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \tag{41e} \]
Vector Norms

- **1-norm** or Taxicab norm or Manhattan norm:
  \[ \| \mathbf{x} \|_1 = \sum_{i=1}^{n} |x_i| \]  
  (42)

- **2-norm** or Euclidean norm or square norm or \(L^2\)-norm:
  \[ \| \mathbf{x} \|_2 = \left( \sum_{i=1}^{n} x^2_i \right)^{1/2} \]  
  (43)

- **\(\infty\)-norm** or infinity-norm or maximum norm:
  \[ \| \mathbf{x} \|_\infty = \max(|x_1|, ..., |x_n|) \]  
  (44)

- Some relations:
  \[ \| \mathbf{x} \|_2 \leq \| \mathbf{x} \|_1 \leq \sqrt{n} \| \mathbf{x} \|_2 \]  
  (45a)
  \[ \| \mathbf{x} \|_\infty \leq \| \mathbf{x} \|_2 \leq \sqrt{n} \| \mathbf{x} \|_\infty \]  
  (45b)
  \[ \| \mathbf{x} \|_\infty \leq \| \mathbf{x} \|_1 \leq n \| \mathbf{x} \|_\infty \]  
  (45c)
Condition Number of a Symmetric Matrix

Definition (Condition Number of a Symmetric Matrix)

The (2-norm) condition number of symmetric (normal) matrix $A$ is:

$$\text{cond}(A) = \|A\|_2\|A^{-1}\|_2 = \frac{\lambda_{\text{max}}(A)}{\lambda_{\text{min}}(A)} \geq 1$$

where $\lambda_{\text{max}}(A)$ is the eigenvalue with the largest magnitude of $A$ and $\lambda_{\text{min}}(A)$ is the smallest.

- In numerical analysis, the condition number of a matrix $A$ is a way of describing how well or badly the system $Ax = b$ could be approximated
  - If $\text{cond}(A)$ is small, the problem is well-conditioned
  - If $\text{cond}(A)$ is large, the problem is rather ill-conditioned
- 📚 Examples and visualizations on Slide 31
Eigenvalue and Condition Number Visualization

\[ \text{Center value is the condition number of } A; \text{ red curve indicates } x^T A x = 1 \]

\[
\begin{align*}
eig(A) &= \begin{bmatrix} 1 & 1 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 1 & 10 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 32 & 0.8 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 5 & 1 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 5 & 1 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 5 & 1 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 5 & 1 \end{bmatrix} \\
eig(A) &= \begin{bmatrix} 5 & 1 \end{bmatrix}
\end{align*}
\]
Jordan Normal Form

- A square matrix $A$ is similar (i.e., $PAP^{-1} = J$) to a block diagonal matrix:

$$J = \text{diag}(J_1, \cdots, J_l)$$  \hspace{1cm} (47)

- Each block $J_i$ is a square matrix where the size of $J_i$ is determined by the algebraic multiplicity of eigenvalue $\lambda_i$ (number of repeated roots in $a(s)$)

- Real eigenvalue $\lambda_i = \alpha$:

$$J_i = \begin{bmatrix}
\alpha & 1 & 0 & \cdots & 0 \\
0 & \alpha & 1 & \ddots & \\
0 & 0 & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \alpha \\
0 & \cdots & 0 & 0 & \alpha \\
\end{bmatrix}$$  \hspace{1cm} (48)

- Complex eigenvalue pair $(\lambda_i, \lambda_{i+1}) = \alpha \pm \beta j$:

$$J_i = \begin{bmatrix}
E_i & I_2 & 0 & \cdots & 0_2 \\
0_2 & E_i & I_2 & \ddots & \\
0_2 & 0_2 & \ddots & \ddots & 0_2 \\
\vdots & \ddots & \ddots & \ddots & 0_2 \\
0_2 & \cdots & 0_2 & 0 & E_i \\
\end{bmatrix}$$  \hspace{1cm} (49a)

where: $E_i = \begin{bmatrix}
\alpha & \beta \\
-\beta & \alpha \\
\end{bmatrix}$, $0_2 = \begin{bmatrix} 0 & 0 \end{bmatrix}$

- This is known as #Jordan normal form

- Examples on Slide 33
Examples of Jordan Normal Form

• An example matrix with only real roots:

\[
A = \begin{bmatrix}
5 & 4 & 2 & 1 \\
0 & 1 & -1 & -1 \\
-1 & -1 & 3 & 0 \\
1 & 1 & -1 & 2
\end{bmatrix} \implies J = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 4 & 1 \\
0 & 0 & 0 & 4
\end{bmatrix}
\]

\[
a(s) = (s - 1)(s - 2)(s - 4)^2
\]

\[
\lambda_1 = 1 \quad \lambda_2 = 2 \quad \lambda_3 = \lambda_4 = 4
\]

• An example matrix with complex roots:

\[
A = \begin{bmatrix}
2 & 9 & 0 & 2 \\
-1 & 2 & 1 & 0 \\
0 & 0 & 3 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix} \implies J = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 \\
0 & 0 & 2 & 3 \\
0 & 0 & -3 & 2
\end{bmatrix}
\]

\[
a(s) = (s + 1)(s - 3)(s^2 - 4s + 13)
\]

\[
\lambda_1 = -1 \quad \lambda_2 = 3 \quad \lambda_3 = \lambda_4^* = 2 + 3j
\]
Inverse of a Matrix

**Definition (Inverse of a Square Matrix)**

The **inverse of a square matrix** $A$ is a matrix $A^{-1}$ such that:

$$A^{-1}A = AA^{-1} = I_n$$  \hfill (52)

where $A^{-1}$ exists only if $A$ is nonsingular (or invertible or nondegenerate) which is equivalent to the following statements:

- $\det(A) \neq 0$
- $A$ has full rank; that is, $\text{rank}(A) = n$
- The number 0 is not an eigenvalue of $A$
- The equation $Ax = 0$ has only the trivial solution $x = 0$
- The equation $Ax = b$ has exactly one solution for each $b$ in $\mathbb{R}^n$
→ Orthogonal Matrix

- An **orthogonal matrix** is a square matrix with real entries whose columns and rows are orthogonal unit vectors (that is, orthonormal vectors)

- Equivalently, a matrix $A$ is orthogonal if its transpose is equal to its inverse:

  $$A^T = A^{-1} \quad (53)$$

  which entails:

  $$A^TA = AA^T = I_n \quad (54)$$

- An orthogonal matrix $A$ is necessarily invertible (with inverse $A^{-1} = A^T$), unitary ($A^{-1} = A^*$), and normal ($A^*A = AA^*$)

- The determinant of any orthogonal matrix is either $+1$ or $-1$

- A special orthogonal matrix is an orthogonal matrix with determinant $+1$

- As a linear transformation, every orthogonal matrix with determinant $+1$ is a pure rotation without reflection, that is, the transformation preserves the orientation of the transformed structure

- ✓ See final row on Slide 31 with rotations applied but condition number does not change
→ Matrix Identities

\[ A(B + C) = AB + AC \]  \hspace{1cm} (55a)

\[ (A + B)^T = A^T + B^T \]  \hspace{1cm} (55b)

\[ (AB)^T = B^T A^T \]  \hspace{1cm} (55c)

\[ (AB)^{-1} = B^{-1} A^{-1} \]  \hspace{1cm} (if individual inverses exist)  \hspace{1cm} (55d)

\[ (A^{-1})^T = (A^T)^{-1} \]  \hspace{1cm} (55e)
References for Linear Algebra

- Some references include:
  - Appendix A in *LNO* by Griva, Nash, and Sofer
  - *Interactive Linear Algebra*\(^1\) by Margalit and Rabinoff (free)
  - *A First Course in Linear Algebra* by Beezer (free)
  - *A Concise Introduction to Linear Algebra* by Schay (CSU available)
  - 3Blue1Brown video series\(^2\)

- Some tools include:
  - Matlab\(^3\)
  - Python numpy.linalg\(^4\)

---

\(^1\) [https://textbooks.math.gatech.edu/ila](https://textbooks.math.gatech.edu/ila) — accompanying website

\(^2\) [www.3blue1brown.com/topics/linear-algebra](http://www.3blue1brown.com/topics/linear-algebra)

\(^3\) [www.mathworks.com/help/matlab/linear-algebra.html](http://www.mathworks.com/help/matlab/linear-algebra.html)

3

Multivariable and Matrix Calculus
Gradient

- Let $f$ be a real-valued function of $n$ variables\(^1\):

$$f(x) = f([x_1 \ x_2 \ \cdots \ x_n]^T) \quad (56)$$

**Definition (Gradient)**

The *gradient* of the function $f$ is the vector of first derivatives of $f$ and is notated as:

$$\nabla f(x) \equiv \frac{\partial f(x)}{\partial x} \equiv \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \quad (57)$$

- [Example on Slide 40](#)

\(^1\) Section B.4 in *LNO*
**Hessian**

**Definition (Hessian)**

*The matrix of second derivatives of* $f$ *is called the Hessian matrix or Hessian and is denoted as* $\nabla^2 f$. *The matrix entries are:*

$$\left[\nabla^2 f(x)\right]_{ij} \equiv \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad (58)$$

*or written directly as a matrix:*

$$\nabla^2 f(x) = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n}
\end{bmatrix} \quad (59)$$

- For functions with continuous second derivatives, it will always be a symmetric matrix, that is, $[\nabla^2 f(x)]_{ij} = [\nabla^2 f(x)]_{ji}$

- Example on Slide 40
Example of the Gradient and Hessian

- Consider the function:

\[ f(x) = 2x_1^4 + 3x_1^2x_2 + 2x_1^3 + 4x_2^2 \]  \hspace{1cm} (60)

- The gradient of this function is:

\[ \nabla f(x) = \begin{bmatrix} 8x_1^3 + 6x_1x_2 + 2x_2^3 \\ 3x_1^2 + 6x_1x_2^2 + 8x_2 \end{bmatrix} \]  \hspace{1cm} (61)

- And the Hessian matrix is:

\[ \nabla^2 f(x) = \begin{bmatrix} 24x_1^2 + 6x_2 & 6x_1 + 6x_2^2 \\ 6x_1 + 6x_2 & 12x_1x_2 + 8 \end{bmatrix} \]  \hspace{1cm} (62)

- At the point \( x_0 = [-2, 3]^T \), these matrices evaluate to:

\[ \nabla f(x_0) = \begin{bmatrix} -46 \\ -72 \end{bmatrix} \quad \nabla^2 f(x_0) = \begin{bmatrix} 114 & 42 \\ 42 & -64 \end{bmatrix} \]  \hspace{1cm} (63)
→ Jacobian

- Consider the following vector-valued function:

\[ f(x) = [f_1(x) \quad f_2(x) \quad \cdots \quad f_m(x)]^\top \]  

(64)

where \( f_i \) are real-valued functions and there are \( m \) functions of \( n \) variables

- Then \( \nabla f(x) \) is the matrix with entries:

\[
[\nabla f(x)]_{ij} = \frac{\partial f_j(x)}{\partial x_i}
\]

(65)

or written directly as a matrix:

\[
\nabla f(x) = \begin{bmatrix}
\frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_1(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n}
\end{bmatrix}
\]

(66)

- The Jacobian of \( f \) at the point \( x \) is defined as \( \nabla^\top f(x) \)

- Example on Slide 42
Example of the Jacobian

• Consider the vector-valued function:

\[
f(x) = \begin{bmatrix}
\sin(x_1) + \cos(x_2) \\
e^{3x_1 + x_2^2} \\
4x_1^3 + 7x_1x_2^2
\end{bmatrix}
\]  \hspace{1cm} (67)

• The Jacobian is equal to:

\[
\nabla^T f(x) = \begin{bmatrix}
\cos(x_1) & -\sin(x_2) \\
3e^{3x_1 + x_2^2} & 2x_2e^{3x_1 + x_2^2} \\
12x_1^2 + 7x_2^2 & 14x_1x_2
\end{bmatrix}
\]  \hspace{1cm} (68)

• At the point \( x_0 = [1, 2]^T \), the Jacobian evaluates to:

\[
\nabla^T f(x_0) = \begin{bmatrix}
\cos(1) & -\sin(2) \\
3e^7 & 4e^7 \\
40 & 28
\end{bmatrix}
\]  \hspace{1cm} (69)
Derivatives of a Product

Definition (Multivariable Derivatives of a Product)

Suppose that:

\[ f(x) = g(x)h(x) \]  \hspace{1cm} (70)

where \( g \) and \( h \) are both continuously differentiable functions of the \( n \)-dimensional vector \( x \). Then the gradient is:

\[ \nabla f(x) = \nabla g(x)h(x) + \nabla h(x)g(x) \]  \hspace{1cm} (71)

and the Hessian is:

\[ \nabla^2 f(x) = \nabla^2 g(x)h(x) + \nabla g(x)\nabla^\top h(x) + \nabla^2 h(x)g(x) + \nabla h(x)\nabla^\top g(x) \]  \hspace{1cm} (72)

- Examples on Slide 44

---

\(^1\) Section B.6 in *LNO*
Two Examples

• Consider a system of linear equations:

\[
Ax = b \implies f(x) = Ax - b = 0
\]  (73)

• Then the Jacobian has a simple form:

\[
\nabla^T f(x) = A
\]  (74)

• Consider the function that can be defined by a product:

\[
f(x) = (a^T x)(b^T x)
\]  (75)

where \((a, b, x)\) are \(n\)-dimensional vectors

• Then, the gradient is:

\[
\nabla f(x) = a(b^T x) + b(a^T x)
\]  (76)

• And the Hessian is:

\[
\nabla^2 f(x) = ab^T + ba^T
\]  (77)
Chain Rule

Definition (Matrix Chain Rule)

Consider the composite function\(^1\):

\[ h(t) = g(x(t)) \]  

(78)

The chain rule states that if \( g \) is continuously differentiable in \( \mathbb{R}^n \), and all \( x_i \) are continuously differentiable in \( \mathbb{R}^m \), then \( h \) is continuously differentiable in \( \mathbb{R}^m \) and:

\[ \nabla h(t) = \nabla x(t) \nabla g(x) \]  

(79)

- Example on Slide 46
- Assuming that \( g \) is a scalar function and if \( g \) and \( x_i \) are twice continuously differentiable, then \( h \) is twice continuously differentiable in \( \mathbb{R}^m \) and:

\[ \nabla^2 h(t) = \nabla^2 x(t) \nabla g(x(t)) + \nabla x(t) \nabla^2 g(x(t)) \nabla^\top x(t) \]  

(80)

\(^1\) Section B.7 in \textit{LNO}
### Example of the Chain Rule

- Let $h(t) = g(x(t))$ with:

  $$g_1(x) = x_1^2 - x_1 x_2 \quad g_2(x) = -x_1^4 + 2x_2^2 \quad x_1(t) = t_1 + 2t_2 - 3t_3 \quad x_2(t) = t_1^2 + t_2$$  \hspace{1cm} (81)

- Then:

  $$\nabla x(t) = \begin{bmatrix} 1 & 2t_1 \\ 2 & 1 \\ -3 & 0 \end{bmatrix} \quad \nabla g(x(t)) = \begin{bmatrix} 2x_1(t) - x_2(t) & -4x_1^3(t) \\ -4x_1(t) & 4x_2(t) \end{bmatrix}$$  \hspace{1cm} (82a)

  $$= \begin{bmatrix} 2(t_1 + 2t_2 - 3t_3) - (t_1^2 + t_2) & -4(t_1 + 2t_2 - 3t_3)^3 \\ -4(t_1 + 2t_2 - 3t_3) & 4(t_1^2 + t_2) \end{bmatrix}$$  \hspace{1cm} (82b)

- Hence:

  $$\nabla h(t) = \begin{bmatrix} 1 & 2t_1 \\ 2 & 1 \\ -3 & 0 \end{bmatrix} \begin{bmatrix} 2(t_1 + 2t_2 - 3t_3) - (t_1^2 + t_2) & -4(t_1 + 2t_2 - 3t_3)^3 \\ -4(t_1 + 2t_2 - 3t_3) & 4(t_1^2 + t_2) \end{bmatrix}$$  \hspace{1cm} (83)
## Table of Identities (1) — Vector-by-Vector

<table>
<thead>
<tr>
<th>Expression</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(x)$</td>
<td>$\nabla f(x)$</td>
</tr>
<tr>
<td>$v(x) \cdot u(x)$</td>
<td>$v(x) \cdot \nabla u(x) + \nabla v(x) \cdot [u(x)]^\top$</td>
</tr>
<tr>
<td>$u(x) + v(x)$</td>
<td>$\nabla u(x) + \nabla v(x)$</td>
</tr>
<tr>
<td>$g(u(x))$</td>
<td>$\nabla u(x) \cdot \nabla g(u)$</td>
</tr>
<tr>
<td>$f(g(u(x)))$</td>
<td>$\nabla u(x) \cdot \nabla g(u) \cdot \nabla f(g)$</td>
</tr>
<tr>
<td>$Au(x)$</td>
<td>$\nabla u(x) \cdot A^\top$</td>
</tr>
<tr>
<td>$a$</td>
<td>$0$</td>
</tr>
<tr>
<td>$x$</td>
<td>$I$</td>
</tr>
<tr>
<td>$Ax$</td>
<td>$A^\top$</td>
</tr>
</tbody>
</table>
### Table of Identities (2) — Scalar-by-Vector

<table>
<thead>
<tr>
<th>$f(x)$</th>
<th>$\nabla f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[v(x)]^T \cdot u(x)$</td>
<td>$\nabla u(x) \cdot v(x) + \nabla v(x) \cdot u(x)$</td>
</tr>
<tr>
<td>$a^T u(x)$</td>
<td>$\nabla u(x) \cdot a$</td>
</tr>
<tr>
<td>$x^T Ax$</td>
<td>$(A + A^T)x$</td>
</tr>
<tr>
<td>$x^T x$</td>
<td>$2x$</td>
</tr>
<tr>
<td>$(Ax + b)^T C(Dx + e)$</td>
<td>$A^T C(Dx + e) + D^T C^T (Ax + b)$</td>
</tr>
<tr>
<td>$|x - a|_1$</td>
<td>$\text{sign}(x - a)$</td>
</tr>
<tr>
<td>$|x - a|_2$</td>
<td>$\frac{x - a}{|x - a|_2}$</td>
</tr>
</tbody>
</table>
References for Multivariable Calculus

- Some references include:
  - Sections B.4–B.7 in *LNO* by Griva, Nash, and Sofer
  - Barnes 2006 (free)
  - Hong n.d. (free)
  - Roweis 1999 (free)
  - *Active Calculus Multivariable*¹ by Schlicker, Austin, and Boelkins (free)

- Some tools include:
  - Online tool for matrix calculus²
  - Matlab³

---

¹ [https://activecalculus.org/multi/](https://activecalculus.org/multi/) — accompanying website
² [www.matrixcalculus.org](http://www.matrixcalculus.org)
Summary
Session 0 Summary

- Calculus is the mathematical study of continuous change and is fundamental to many scientific disciplines, including physics, engineering, and economics.
- Linear algebra is the branch of mathematics concerning linear equations and also is fundamental to many of the same disciplines as it is essential to defining and understanding practical and robustly computable problems.
- Matrix calculus is a specialized notation for doing multivariable calculus, especially over spaces of matrices.
→ Terms

# derivative is on Slide 6
# matrix is on Slide 16
# transposed matrix is on Slide 17
# symmetric matrix is on Slide 18
# matrix product is on Slide 19
# diagonal matrix is on Slide 20
# block diagonal matrix is on Slide 21
# trace is on Slide 22
# determinant is on Slide 23
# rank is on Slide 25
# identity matrix is on Slide 26
# characteristic polynomial is on Slide 27
# eigenvalues is on Slide 27
# eigenvector is on Slide 27
# 1-norm is on Slide 29
→ Terms (Continued)

- `2-norm` is on Slide 29
- `∞-norm` is on Slide 29
- `condition number` is on Slide 30
- `Jordan normal form` is on Slide 32
- `inverse of a square matrix` is on Slide 34
- `orthogonal matrix` is on Slide 35
- `gradient` is on Slide 38
- `Hessian matrix or Hessian` is on Slide 39
- `vector-valued function` is on Slide 41
- `Jacobian` is on Slide 41
References

- — (n.d.). *Interactive Linear Algebra.* Online. URL: https://textbooks.math.gatech.edu/ila
- S. Schlicker (n.d.). *Active Calculus - Multivariable.* Online. URL: https://activecalculus.org/multi/
References (Continued)

S. Schlicker, D. Austin, and M. Boelkins (2018). *Active Calculus Multivariable*. Grand Valley State University Libraries. URL: https://scholarworks.gvsu.edu/books/19/