Continuous effort - not strength or intelligence - is the key to unlocking our potential. *W. Churchill*
Before we get started…

- **Last Time:**
  - Class Intro + Syllabus Outline
    - Final Project discussion
    - Established time/date of midterm exam
    - Trip to Iowa and John Deere looks likely
  - Started review on Linear Algebra

- **Today:**
  - Finish review of Linear Algebra
  - Review of Calculus (two definitions and three theorems)

- **HW:** posted on class website, due on Jan. 28.
Matrix Review [Cntd.]

- Symmetric matrix: a square matrix \( A \) for which \( A = A^T \)
- Skew-symmetric matrix: a square matrix \( B \) for which \( B = -B^T \)
- Examples:
  \[
  A = \begin{bmatrix}
  2 & 1 & -1 \\
  1 & 0 & 3 \\
  -1 & 3 & 4
  \end{bmatrix}, \quad
  B = \begin{bmatrix}
  0 & -1 & 2 \\
  1 & 0 & 4 \\
  -2 & -4 & 0
  \end{bmatrix}
  \]

- Singular matrix: square matrix whose determinant is zero
  \[
  \det(A) = 0, \quad A \in \mathbb{R}^{n \times n}
  \]

- Inverse of a square matrix \( A \): a matrix of the same dimension, called \( A^{-1} \), that satisfies the following:
  \[
  A^{-1} \cdot A = A \cdot A^{-1} = I_n, \quad A \in \mathbb{R}^{n \times n}
  \]
Singular vs. Nonsingular Matrices

Let $A$ be a square matrix of dimension $n$. The following are equivalent:

- $Ax = b$ has a unique solution for any $b \in \mathbb{R}^n$.
- $Ax = b$ has a solution for any $b \in \mathbb{R}^n$.
- $Ax = 0$ implies $x = 0$.
- $A^{-1}$ exists.
- $\text{Determinant}(A) \neq 0$.
- $\text{rank}(A) = n$. 
Orthogonal & Orthonormal Matrices

- Definition ($Q$, orthogonal matrix): a square matrix $Q$ is orthogonal if the product $Q^TQ$ is a diagonal matrix.

- Matrix $Q$ is called orthonormal if it’s orthogonal and also $Q^TQ = I_n$.
  - Note that people in general don’t make a distinction between an orthogonal and orthonormal matrix.

- Note that if $Q$ is an orthonormal matrix, then $Q^{-1} = Q^T$.

- Example, orthonormal matrix:

$$Q = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & -1
\end{bmatrix}$$
Remark:

On the Columns of an Orthonormal Matrix

- Assume $Q$ is an orthonormal matrix

$$Q \in \mathbb{R}^{n \times n} \quad Q = [q_1, \ldots, q_n] \quad \text{← orthnormal}$$

$$Q^T Q = I \quad \Rightarrow \quad \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} [q_1, \ldots, q_n] = \begin{bmatrix} q_1^T q_1 & \cdots & q_1^T q_n \\ \vdots & \ddots & \vdots \\ q_n^T q_1 & \cdots & q_n^T q_n \end{bmatrix}$$

$$q_i^T \cdot q_j = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- In other words, the columns (and the rows) of an orthonormal matrix have unit norm and are mutually perpendicular to each other
Condition Number of a Matrix

- Let $A$ be a square matrix. By definition, its condition number is

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$$

- Note that condition number depends on the norm used in its evaluation

- The concept of ill-conditioned linear system $Ax=b$:
  - A system for which small perturbations in $b$ lead to large changes in solution $x$
  - NOTE: A linear system is ill-condition if $\text{cond}(A)$ is large

- Three quick remarks:
  - The closer a matrix is to being singular, the larger its condition number
  - You can’t get $\text{cond}(A)$ to be smaller than 1
  - If $Q$ is orthonormal, then $\text{cond}(Q)=1$
Condition Number of a Matrix

Example

\[
\begin{align*}
7x_1 + 10x_2 &= b_1 \\
5x_1 + 7x_2 &= b_2
\end{align*}
\]

\[
A = \begin{bmatrix}
7 & 10 \\
5 & 7
\end{bmatrix} \quad A^{-1} = \begin{bmatrix}
-7 & 10 \\
5 & -7
\end{bmatrix}
\]

\[
\text{cond}(A)_1 = \|A\|_1 \cdot \|A^{-1}\|_1 = 289
\]

\[
\text{cond}(A)_2 = \|A\|_2 \cdot \|A^{-1}\|_2 \approx 223
\]

\[
\text{cond}(A)_\infty = \|A\|_\infty \cdot \|A^{-1}\|_\infty = 289
\]
Other Useful Formulas

- If $\mathbf{A}$ and $\mathbf{B}$ are invertible, their product is invertible and

  $$(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$$

- Also,

  $$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$$

- For any two matrices $\mathbf{A}$ and $\mathbf{B}$ that can be multiplied

  $$\left(\mathbf{AB}\right)^T = \mathbf{B}^T \mathbf{A}^T$$

- For any three matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ that can be multiplied

  $$\left(\mathbf{AB}\right)\mathbf{C} = \mathbf{A} \left(\mathbf{BC}\right)$$
Lagrange Multiplier Theorem

Theorem:

Assume that a vector \( \mathbf{b} \in \mathbb{R}^n \), and a matrix \( \mathbf{A} \in \mathbb{R}^{m \times n} \), with \( m < n \), are such that for ANY vector \( \mathbf{x} \in \mathbb{R}^n \), one has that \( \mathbf{x}^T \mathbf{b} = 0 \) as soon as \( \mathbf{A} \mathbf{x} = 0 \).

Then it turns out that there is a relationship between \( \mathbf{A} \) and \( \mathbf{b} \), and in fact \( \mathbf{b} \) is a linear combination of the rows of \( \mathbf{A} \). In other words, there is a so called “Lagrange Multiplier” \( \lambda \) such that \( \mathbf{b} = -\mathbf{A}^T \lambda \), or equivalently, \( \mathbf{b} + \mathbf{A}^T \lambda = 0 \).
Lagrange Multiplier Theorem

- Theorem:

Assume that a vector $b \in \mathbb{R}^n$, and a matrix $A \in \mathbb{R}^{m \times n}$, with $m < n$, are two quantities related by the following relationship: **ANY** vector $x \in \mathbb{R}^n$ that is perpendicular on the rows on $A$ is also perpendicular on $b$; i.e., $x^T b = 0$ as soon as $Ax = 0$.

Then it turns out that in fact $b$ is a linear combination of the rows of $A$. In other words, there is a so called “Lagrange Multiplier” $\lambda$ such that $b = -A^T \lambda$, or equivalently, $b + A^T \lambda = 0$. 
Example: Lagrange Multipliers

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

\[ b = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \]

- First, show that any for any \( x = [x_1 \ x_2 \ x_3]^T \), one has that \( x^T b = 0 \) as soon as \( Ax = 0 \)

- Next, show that there is indeed a vector \( \lambda \) such that \( b + A^T \lambda = 0 \)
End: Review of Linear Algebra
Begin: Review of Calculus
Derivatives of Functions

GOAL: Understand how to

- Take **time derivatives** of vectors and matrices

- Take **partial derivatives** of a function with respect to its arguments
  - We will use a matrix-vector notation for computing these partial derivs.
  - Taking partial derivatives might be challenging in the beginning
  - The use of partial derivatives is a recurring theme in the literature
Taking time derivatives of a time dependent vector

FRAMEWORK:

- Vector \( \mathbf{r} \) is represented as a function of time, and it has three components: \( x(t) \), \( y(t) \), \( z(t) \):
  \[
  \mathbf{r}(t) = \begin{bmatrix}
  x(t) \\
  y(t) \\
  z(t)
  \end{bmatrix}
  \]

- Its components change, but the vector is represented in a fixed reference frame

THEN:

\[
\dot{\mathbf{r}}(t) = \begin{bmatrix}
  \dot{x}(t) \\
  \dot{y}(t) \\
  \dot{z}(t)
  \end{bmatrix}, \quad \ddot{\mathbf{r}}(t) = \begin{bmatrix}
  \ddot{x}(t) \\
  \ddot{y}(t) \\
  \ddot{z}(t)
  \end{bmatrix}, \quad \text{etc.}
\]
Time Derivatives, Vector Related Operations

- Assume that $\alpha \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^n$ depend on time. Then it can be proved that the following hold:

\[
\frac{d}{dt}(\alpha \mathbf{a}) = \frac{d\alpha}{dt} \mathbf{a} + \alpha \frac{d\mathbf{a}}{dt} = \dot{\alpha} \mathbf{a} + \alpha \dot{\mathbf{a}}
\]

\[
\frac{d}{dt}(\mathbf{a} + \mathbf{b}) = \frac{d\mathbf{a}}{dt} + \frac{d\mathbf{b}}{dt} = \dot{\mathbf{a}} + \dot{\mathbf{b}}
\]

\[
\frac{d}{dt}(\mathbf{a}^T \mathbf{b}) = \frac{d\mathbf{a}^T}{dt} \mathbf{b} + \mathbf{a}^T \frac{d\mathbf{b}}{dt} = \dot{\mathbf{a}}^T \mathbf{b} + \mathbf{a}^T \dot{\mathbf{b}}
\]

\[
\mathbf{a}^T \mathbf{a} = \text{const} \quad \Rightarrow \quad \mathbf{a}^T \dot{\mathbf{a}} = 0
\]
Taking time derivatives of MATRICES

- By **definition**, the time derivative of a matrix is obtained by taking the time derivative of each entry in the matrix.

- A simple extension of what we’ve seen for vector derivatives.

- Assume that $\alpha \in \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{n \times p}$ depend on time. Then it can be proved that the following hold:

\[
\frac{d}{dt}(\alpha A) = \frac{d\alpha}{dt} A + \alpha \frac{dA}{dt} = \dot{\alpha} A + \alpha \dot{A},
\]

\[
\frac{d}{dt}(A + B) = \frac{dA}{dt} + \frac{dB}{dt} = \dot{A} + \dot{B},
\]

\[
\frac{d}{dt}(AC) = \frac{dA}{dt} C + A \frac{dC}{dt} = \dot{A} C + A \dot{C}.
\]
Done with Time Derivatives

... Moving on to Partial Derivatives
Derivatives of Functions: Why Bother?

- Partial derivatives are essential in this class
  - In computing the Jacobian matrix associated with the constraints that define the joints present in a mechanism
  - Essential in computing the Jacobian matrix of any nonlinear system that you will have to solve when using implicit integration to find the time evolution of a dynamic system

- Beyond this class
  - Whenever you do a sensitivity analysis (in optimization, for instance) you need partial derivatives of your functions
What’s the story behind the concept of partial derivative?

- What’s the meaning of a partial derivative?
  - It captures the “sensitivity” of a function quantity with respect to a variable the function depends upon
  - Shows how much the function changes when the variable changes a bit

- Simplest case of partial derivative: you have one function that depends on one variable:

\[ f(x) = \ln x \quad , \quad g(z) = \sin(4z + \pi) \quad , \quad \text{etc.} \]

- Then,

\[ \frac{\partial f}{\partial x} = \frac{1}{x} \quad , \quad \frac{\partial g}{\partial z} = 4 \cos(4z + \pi) \quad , \quad \text{etc.} \]
Partial Derivative, Two Variables

- Suppose you have one function but it depends on **two** variables, say x and y:

  \[ f(x, y) = \sin(x^2 + 3y^2) \]

- To simplify the notation, an array \( \mathbf{q} \) is introduced:

  \[ \mathbf{q} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \]

- With this, the partial derivative of \( f(\mathbf{q}) \) wrt \( \mathbf{q} \) is **defined** as

  \[ \frac{\partial f}{\partial \mathbf{q}} = f_{\mathbf{q}} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = [2x \cos(x^2 + 3y^2) \quad 6y \cos(x^2 + 3y^2)] \]
...and here is as good as it gets (vector function)

- You have a group of “m” functions that are gathered together in an array, and they depend on a collection of “n” variables:

\[ f_1, f_2, \ldots, f_m \text{ depend on } x_1, x_2, \ldots, x_n \]

- The array that collects all “m” functions is called \( \mathbf{F} \):

\[
\mathbf{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix}
    f_1(x_1, x_2, \ldots, x_n) \\
    f_2(x_1, x_2, \ldots, x_n) \\
    \vdots \\
    f_m(x_1, x_2, \ldots, x_n)
\end{bmatrix} \in \mathbb{R}^m
\]

- The array that collects all “n” variables is called \( \mathbf{q} \):

\[
\mathbf{q} = \begin{bmatrix}
    x_1 \\
    x_2 \\
    \ldots \\
    x_n
\end{bmatrix} \in \mathbb{R}^n
\]
Most general partial derivative
(Vector Function, cntd)

● Then, in the most general case, by definition

\[
\frac{\partial \mathbf{F}}{\partial \mathbf{q}} = \mathbf{F}_q = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n}
\end{bmatrix} \in \mathbb{R}^{m \times n}
\]

● Example 2.5.2:

\[
\mathbf{q} = \begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix} \quad \mathbf{r}^P = \begin{bmatrix}
\cos \theta_1 + l \cos(\theta_1 + \theta_2) \\
\sin \theta_1 + l \sin(\theta_1 + \theta_2)
\end{bmatrix} \quad \mathbf{r}_q^P = ?
\]
Example: Left and Right mean the same thing

- Let $x$, $y$, and $\phi$ be three generalized coordinates.

- Define the function $r$ of $x$, $y$, and $\phi$ as

$$r(x, y, \phi) = \left[ \begin{array}{c} x + 2l\cos \phi \\ y - 2l\sin \phi \end{array} \right]$$

- Compute the partial derivatives

$$r_{x,y,\phi} = \left[ \begin{array}{ccc} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} & \frac{\partial r}{\partial \phi} \end{array} \right]$$

- Let $x$, $y$, and $\phi$ be three generalized coordinates, and define the array $\mathbf{q}$

$$\mathbf{q} = \left[ \begin{array}{c} x \\ y \\ \phi \end{array} \right]$$

- Define the function $r$ of $\mathbf{q}$:

$$r(\mathbf{q}) = \left[ \begin{array}{c} x + 2l\cos \phi \\ y - 2l\sin \phi \end{array} \right]$$

- Compute the partial derivative

$$r_{\mathbf{q}} = \frac{\partial r}{\partial \mathbf{q}}$$
Exercise

\[ q = \begin{bmatrix} x \\ y \\ \phi \end{bmatrix} \]

\[ r(q) = \begin{bmatrix} x + 2l \cos \phi \\ y - 2l \sin \phi \end{bmatrix} \]

\[ r_q = \frac{\partial r}{\partial q} = ? \]

\[ r_q = \frac{\partial r}{\partial q} = \begin{bmatrix} \frac{\partial r}{\partial q_1} & \frac{\partial r}{\partial q_2} & \frac{\partial r}{\partial q_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} & \frac{\partial r}{\partial \phi} \end{bmatrix} \]

\[ \downarrow \]

\[ r_q = \begin{bmatrix} 1 & 0 & -2l \sin \varphi \\ 0 & 1 & -2l \cos \varphi \end{bmatrix} \]
Partial Derivatives: Good to Remember...

- In the most general case, you start with “m” functions in “n” variables, and end with an \((m \times n)\) matrix of partial derivatives.
  - You start with a column vector of functions and then end up with a matrix.

- Taking a partial derivative leads to a *higher dimension* quantity:
  - Scalar Function – leads to row vector
  - Vector Function – leads to matrix
  - I call this the “accordion rule.”

- In this class, taking partial derivatives can lead to one of the following:
  - A row vector
  - A full blown matrix
  - If you see something else chances are you made a mistake…

- So far, we only introduced a couple of *definitions*. 
Done with Partial Derivatives

…

Moving on to Chain Rule of Differentiation
Scenario 1: **Scalar Function**

- $f$ is a function of “$n$” variables: $q_1, \ldots, q_n$
  \[ f : \mathbb{R}^n \to \mathbb{R} \]

- However, each of these variables $q_i$ in turn depends on a set of “$k$” other variables $x_1, \ldots, x_k$.

  \[
  q = \begin{bmatrix}
  q_1(x_1, \ldots, x_k) \\
  \vdots \\
  q_n(x_1, \ldots, x_k)
  \end{bmatrix} : \mathbb{R}^k \to \mathbb{R}^n
  \]

- The composition of $f$ and $q$ leads to a new function $\phi(x)$:

  \[
  \phi(x) = f \circ q = f(q(x)) : \mathbb{R}^k \to \mathbb{R}
  \]
Chain Rule for a **Scalar Function**

- The question: how do you compute \( \phi_x \)?
  - Using our notation:

\[
\phi = f \circ q = f(q(x)) \quad \Rightarrow \quad \phi_x = \frac{\partial \phi}{\partial x} = ??
\]

- **Theorem**: Chain rule of differentiation for scalar function

\[
\phi_x = \frac{\partial \phi}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x} = f_q \cdot q_x
\]

(This theorem is proved in your elementary calculus class)
Example

Assume that \( \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \) and a function \( \phi \) of \( \mathbf{y} \) is defined as: \( \phi(\mathbf{y}) = 3y_1^2 + \sin y_2 \).

In turn, \( \mathbf{y} \) depends on a variable \( \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \) as follows:

\[
\mathbf{y} = \mathbf{y}(\mathbf{x}) = \begin{bmatrix} y_1(\mathbf{x}) \\ y_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 2x_1 + \log x_2 + \sqrt{x_3} \\ (x_1 - x_2)^2 \end{bmatrix}
\]

Now, since \( \phi \) depends on \( \mathbf{y} \) and \( \mathbf{y} \) depends on \( \mathbf{x} \), it means that \( \phi \) depends on \( \mathbf{x} \). Find the partial derivative of \( \phi \) with respect to \( \mathbf{x} \), that is,

\[
\phi_{\mathbf{x}} = \frac{\partial \phi}{\partial \mathbf{x}} = \left[ \frac{\partial \phi}{\partial x_1} \quad \frac{\partial \phi}{\partial x_2} \quad \frac{\partial \phi}{\partial x_3} \right] = ?
\]
Scenario 2: Vector Function

- **F** is a function of “n” variables: $q_1, \ldots, q_n$
  \[ F : \mathbb{R}^n \rightarrow \mathbb{R}^m \]

- However, each of these variables $q_i$ in turn depends on a set of “k” other variables $x_1, \ldots, x_k$.
  \[ q = \begin{bmatrix} q_1(x_1, \ldots, x_k) \\ \vdots \\ q_n(x_1, \ldots, x_k) \end{bmatrix} : \mathbb{R}^k \rightarrow \mathbb{R}^n \]

- The composition of **F** and **q** leads to a new function $\Phi(x)$:
  \[ \Phi(x) = F \circ q = F(q(x)) : \mathbb{R}^k \rightarrow \mathbb{R}^m \]
Chain Rule for a **Vector** Function

- How do you compute the partial derivative of $\Phi$?

  \[ \Phi : \mathbb{R}^k \rightarrow \mathbb{R}^m \]

  \[ \Phi = \Phi(q(x)) \quad \Rightarrow \quad \Phi_x = \frac{\partial \Phi}{\partial x} = ?? \]

- **Theorem**: Chain rule of differentiation for vector functions

  \[ \Phi_x = \frac{\partial \Phi}{\partial x} = \frac{\partial F}{\partial q} \cdot \frac{\partial q}{\partial x} \]

(This theorem is proved in your elementary calculus class)
Example

Assume that \( y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \) and a function \( f \) of \( y \) is defined as: \( f(y) = \begin{bmatrix} 2y_1 + y_2^2 \\ y_1y_2 \end{bmatrix} \).

In turn, \( y \) depends on a variable \( x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \) as follows:

\[
y = y(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} x_1x_2 \\ x_1^2 - x_2 \end{bmatrix}
\]

Now, since \( f \) depends on \( y \) and \( y \) depends on \( x \), it means that \( f \) depends on \( x \). Find the partial derivative of \( f \) with respect to \( x \), that is,

\[
f_x = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \ ?
\]