ELEC6229
Advanced Systems and Signal Processing
Linear Algebra and Probability

Bing Chu
Electronics and Computer Science
University of Southampton
Email: b.chu@ecs.soton.ac.uk
Office: Building 1, Room 2007
By the end of this week’s lectures, you should be able to

- Use matrix properties and basic probability theory to explain and solve various problems

Reading: Chapters 1, 2 Dan Simon’s book
Lecture Overview

- Vectors and Matrices
- Determinant and Matrix Inverse
- Positive Definite Matrix and SVD
- Matrix and Vector Calculus
Vectors and Matrices

- **Scalar**: $k \in R$ or $C$
- **Vector**:
  - Row vector: $x \in R^n$ or $C^n$
  - Column Vector: $x \in R^n$ or $C^n$
- **Operation**:
  - Scalar multiplication
  - Vector addition
Vectors and Matrices

- **Matrix**: \( A \in \mathbb{R}^{n \times m} \) or \( C^{n \times m} \)
  - Vector as a special case of matrix

- **Special matrices**:
  - identity matrix, zero matrix, diagonal matrix,
  - triangular matrix, Toeplitz matrix, Hankel matrix ...

- **Matrix operation**
  - \( kA, A^T, A^H \)
Matrix Algebra

- **Basic Operation:**
  - $A + B$
  - $A \times B$, $B \times A$
  - $(AB)^T = B^T A^T$

- **Block matrices:**
  - $A + B$, $A \times B$
  - Simplify big matrix calculation
Rank of a Matrix

- **Rank**: $A \in R^{m \times n}, B \in R^{n \times k}$
  - $\text{rank}(A) \leq \text{min}(m, n)$

- **Sylvester’s rank inequality**:
  
  $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \text{min}(\text{rank}(A), \text{rank}(B))$
Square Matrix: $A \in \mathbb{R}^{n \times n}$

- **Trace:** $tr(A)$
  - $tr(A) = tr(A^T)$
  - $tr(A + B) = tr(A) + tr(B)$

- **Exercise:**
  - Show $tr(AB) = tr(BA)$
  - When $A = x, B = x^T$?
Square Matrix: eigenvalues

- Eigenvalue: \( Ax = \lambda x, \ x \neq 0 \)
  - \( A^k \Rightarrow \lambda^k \)
  - \( A + \alpha I \Rightarrow \lambda + \alpha \)
  - \( tr(A) = \sum_{i=1}^{n} \lambda_i \)

- Diagonalisation of a square matrix:
  - Jordon canonical form
Square Matrix: determinant

- **Determinant**: $det(A)$ or $|A|$, Laplace expansion
  \[ |A| = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} |A^{(i,j)}| \]

- **Properties**:
  - $|AB| = |A||B|$, $|\alpha A| = \alpha^n |A|$, $|A| = \prod_{i=1}^{n} \lambda_i$
  - $A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{n \times p}$
    \[ |I_p + AB| = |I_n + BA| \]
Square Matrix: Inverse

- Inverse matrix $A^{-1} : A \in R^{n \times n}$
  \[ AA^{-1} = A^{-1}A = I \]

- Properties:
  - $A^{-1}$ exists $\iff |A| \neq 0 \iff \text{rank}(A) = n \iff 0$ is not an eigenvalue
  - $(AB)^{-1} = B^{-1}A^{-1}$
Matrix Inversion Lemma

\[
(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}
\]

- **Properties:**
  - **Consider**

\[
A = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\]

\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix} = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|
\]

\[
= |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|
\]
Symmetric Matrix

- **Symmetric matrix** $A$:
  \[ A = A^T \in \mathbb{R}^{n \times n} \]

- **Properties:**
  - *All eigenvalues are real*
  - *Orthogonal diagonalisation*
    \[ Q^T AQ = \text{diag}\{\cdots\} \]
    
    *where* \( Q^T Q = I \)
Positive Definite Matrix

- Positive definite matrix $A > 0$: $A = A^T \in \mathbb{R}^{n \times n}$
  
  $x^T Ax > 0, \forall x \neq 0$

- Properties:
  - $\text{eig}(A) > 0$
  - $A > 0 \Rightarrow A^{-1} \text{ exists and } A^{-1} > 0$
  - $A > 0 \text{ and } |F| \neq 0 \Rightarrow F^T AF > 0$
  - $A > 0 \text{ and } B > 0 \Rightarrow A + B > 0$
  - $A > 0 \text{ and } B > 0, A > B \text{ means } A - B > 0$
Singular Value Decomposition (SVD)

- Singular value decomposition: \( A \in \mathbb{R}^{m \times n} \)
  \[
  A = U \Sigma V^T
  \]
  where \( U \) and \( V \) are orthogonal (unitary) matrices, and \( \Sigma \in \mathbb{R}^{m \times n} \) rectangular diagonal matrix with nonnegative diagonal elements (\( \sigma_i \)).

- Properties:
  - \( \sigma_i(A) = \sqrt{\lambda_i(A^T A)} = \sqrt{\lambda_i(A A^T)} \)
  - Note: diagonalisation and SVD for square matrix
We consider function $f$

<table>
<thead>
<tr>
<th>Domain</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>$R$</td>
</tr>
<tr>
<td>$\mathbb{R}^n$</td>
<td>$\mathbb{R}^n$</td>
</tr>
<tr>
<td>$\mathbb{R}^{m \times n}$</td>
<td>$\mathbb{R}^{m \times n}$</td>
</tr>
</tbody>
</table>
Matrix Calculus: $R \rightarrow R^{m \times n}$

- **Element wise operation:** $A(t) \in R^{m \times n}$

  \[
  \frac{dA(t)}{dt} = \begin{bmatrix}
  \frac{dA_{11}(t)}{dt} & \ldots & \frac{dA_{1n}(t)}{dt} \\
  \vdots & \ddots & \vdots \\
  \frac{dA_{m1}(t)}{dt} & \ldots & \frac{dA_{mn}(t)}{dt}
  \end{bmatrix}
  \]

- **Property:** $A(t) \in R^{n \times n}$

  \[
  \frac{dA^{-1}(t)}{dt} = -A^{-1}(t) \frac{dA(t)}{dt} A^{-1}(t)
  \]
Matrix Calculus: $\mathbb{R}^n \rightarrow \mathbb{R}$

- Two different definitions:
  \[
  \frac{\partial f}{\partial x} = \begin{bmatrix}
    \frac{\partial f}{\partial x_1} & \ldots & \frac{\partial f}{\partial x_n}
  \end{bmatrix}
  \]

- Properties:
  - $y \in \mathbb{R}^n$, $\frac{\partial x^T y}{\partial x} = y^T$
  - $A \in \mathbb{R}^{n \times n}$, $\frac{\partial x^T A x}{\partial x} = x^T A + x^T A^T$
Matrix Calculus: $\mathbb{R}^n \rightarrow \mathbb{R}^m$

- **Two different definition:**

\[
\frac{\partial g}{\partial x} = \begin{bmatrix}
\frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n}
\end{bmatrix}
\]

- **Properties:**

\[
\frac{\partial Ax}{\partial x} = A, \quad \frac{\partial x^T A}{\partial x} = A^T
\]
Matrix Calculus: Identities

\[
\frac{\partial (u + v)}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}
\]

\[
\frac{\partial (u \times v)}{\partial x} = u \times \frac{\partial v}{\partial x} + \frac{\partial u}{\partial x} \times v
\]

\[
\frac{\partial g(u)}{\partial x} = \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial x}
\]
Part B

Probability
Let’s Make a Deal
Lecture Overview

- Probability
- Random Variable
- Multiple Random Variables
- Stochastic Process
Probability

- Probability of an event: $P(A)$
- Conditional probability:

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

- Bayes' Rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$
Random Variable (RV)

- Random variable: $X$
- PDF and pdf:
  - PDF: $F_X(x) = P(X \leq x)$
  - pdf: $f_X(x) = \frac{dF_X(x)}{dx}$
- Special random variables
  - Uniform RV
  - Gaussian RV
Bayes' Rule:

\[ f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)}{f_{X_2}(x_2)} \]

Chapman-Kolmogorov equation:

\[ f[x_1|(x_2, x_3, x_4)]f[(x_2, x_3)|x_4] = f[(x_1, x_2, x_3)|x_4)] \]
Moment

- $i^{th}$ moment of $X$: $E(X^i)$
- $i^{th}$ central moment of $X$: $E[(X - \bar{x})^i]$
- Special cases:
  - $i = 1$
  - $i = 2$
  - $i = 3$
Function of Random Variable

- **Random variable:** $X$
- **Function of $X$:** $Y = g(X)$

$$f_Y(y) = \sum_i f_X[g_i^{-1}(y)] \left| \frac{dg_i^{-1}(y)}{dy} \right|$$

- **Special cases:**
  - **Linear function:** $Y = aX + b$
Multiple Random Variables

- Two RVs: \( X, Y \)
- Joint PDF and pdf:
  - PDF: \( F_{XY}(x, y) = P(X \leq x, Y \leq y) \)
  - pdf: \( f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} \)
- Function of two RVs:
  \[
  E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \, dx \, dy
  \]
Statistic Independence

- **Independence:**
  \[ P(X \leq x, Y \leq y) = P(X \leq x)P(y \leq y) \]

- **Correlation:** \( R_{XY} = E(XY) \)
  - **Correlation coefficient:**
    \[ \rho = \frac{E[(X-\bar{x})(Y-\bar{y})]}{\sqrt{E[(X-\bar{x})^2]}\sqrt{E[(Y-\bar{y})^2]}} \]
  - **Correlation and independence?**
Correlation:

\[ R_{XY} = E(XY^T) \]

- Autocorrelation: \( R_X = E(XX^T) \) (PSD)

Covariance:

\[ C_{XY} = E[(X - \bar{x})(Y - \bar{y})^T] \]

- Autocovariance: \( C_X = E[(X - \bar{x})(X - \bar{x})^T] \) (PSD)
A random variable changes with $t$: $X(t)$

- Process/Sequence
- Continuous/Discrete

PDF and pdf:

- PDF: $F_X(x, t) = P(X(t) \leq x)$ (and high order)
- pdf: $f_X(x, t) = \frac{dF_X(x,t)}{dx}$ (and high order)

Mean and variance

Q: how to characterise a stochastic process
Stationary Stochastic Process

- **Strict Sense Stationary (SSS):**
  \[ f_X(x, t) = f_X(x) \text{ (and for high order pdf)} \]

- **Wide Sense Stationary (WSS):**
  - \[ E[X(t)] = \bar{x} \]
  - \[ E[X(t_1)X^T(t_2)] = R_X(t_2 - t_1) \]

- **Properties of** \( R_X \)

- **Ergodic process**
White Noise

- Wiener-Khintchine relation:
  \[ S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega \tau} d\tau \]
  \[ R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega \tau} d\omega \]

- White noise: \( S_X(\omega) = A \quad (R_X(\tau) = A\delta(\tau)) \)

- pdf of white noise?
Let’s Make a Deal
Let’s Make a Deal – Assumptions

- The host must always open a door that was not picked by the contestant.
- The host must always open a door to reveal a goat and never the car.
- The host must always offer the chance to switch between the originally chosen door and the remaining closed door.
Let’s Make a Deal – Bayes’ Calculation

- $C_i$ indicates the car is behind door $i$
- $X_i$ indicates the player initially chooses door $i$
- $H_i$ indicates the host opens door $i$
- **Question**: if players initially select door $i$ and the host opens door $k$, the conditional probability of winning by switching is?
Let’s Make a Deal – Bayes’ Calculation

\[ P(C_2|H_3, X_1) = \frac{P(H_3|C_2, X_1)P(C_2|X_1)}{P(H_3|X_1)} = ? \]

\[ P(C_2|H_3, X_1) = \frac{2}{3} \]
Let’s Make a Deal – New Assumptions

- The host must always open a door that was not picked by the contestant.
- The host must always open a door to reveal a goat and never the car randomly.
- The host must always offer the chance to switch between the originally chosen door and the remaining closed door.
Let’s Make a Deal – Bayes’ Calculation

\[ P(C_2|H_3, X_1) = \frac{P(H_3|C_2, X_1)P(C_2|X_1)}{P(H_3|X_1)} =? \]

\[ P(C_2|H_3, X_1) =? \]
By the end of this week’s lectures, you should be able to

- Use matrix properties and basic probability theory to explain and solve various problems

Reading: Chapters 1, 2 Dan Simon’s book