The course provides an introduction to the mathematics of data analysis and a detailed overview of statistical models for inference and prediction.

**Course Structure:**
- Weekly lectures [notes: www.ece.udel.edu/~arce/Courses]
- Homework & computer assignments [30%]
- Midterm & Final examinations [70%]

**Textbooks:**
- Papoulis and Pillai, Probability, random variables, and stochastic processes.
- Hastie, Tibshirani and Friedman, The elements of statistical learning.
- Haykin, Adaptive Filter Theory.
Objective: Utilize tools from linear algebra to characterize and analyze matrices, especially the correlation matrix

- The correlation matrix plays a large role in statistical characterization and processing.
- Previously result: $\mathbf{R}$ is Hermitian.
- Further insight into the correlation matrix is achieved through eigen analysis
  - Eigenvalues and vectors
  - Matrix diagonalization
  - Application: Optimum filtering problems
**Objective:** For a Hermitian matrix $\mathbf{R}$, find a vector $\mathbf{q}$ satisfying

$$
\mathbf{R}\mathbf{q} = \lambda \mathbf{q}
$$

- **Interpretation:** Linear transformation by $\mathbf{R}$ changes the scale, but not the direction of $\mathbf{q}$
- **Fact:** A $M \times M$ matrix $\mathbf{R}$ has $M$ eigenvectors and eigenvalues

$$
\mathbf{R}\mathbf{q}_i = \lambda_i \mathbf{q}_i \quad i = 1, 2, 3, \ldots, M
$$

To see this, note

$$(\mathbf{R} - \lambda \mathbf{I})\mathbf{q} = \mathbf{0}$$

For this to be true, the row/columns of $(\mathbf{R} - \lambda \mathbf{I})$ must be linearly dependent,

$$\Rightarrow \det(\mathbf{R} - \lambda \mathbf{I}) = 0$$
**Note:** \( \text{det}(R - \lambda I) \) is a \( M \)th order polynomial in \( \lambda \)

- The roots of the polynomial are the eigenvalues \( \lambda_1, \lambda_2, \cdots, \lambda_M \)

\[
Rq_i = \lambda_i q_i
\]

- Each eigenvector \( q_i \) is associated with one eigenvalue \( \lambda_i \)
- The eigenvectors are not unique

\[
Rq_i = \lambda_i q_i \\
\Rightarrow R(aq_i) = \lambda_i (aq_i)
\]

**Consequence:** eigenvectors are generally normalized, e.g., \( |q_i| = 1 \) for \( i = 1, 2, \ldots, M \)
Example (General two dimensional case)

Let $M = 2$ and

$$
R = \begin{bmatrix}
R_{1,1} & R_{1,2} \\
R_{2,1} & R_{2,2}
\end{bmatrix}
$$

Determine the eigenvalues and eigenvectors.

Thus

$$
\det(R - \lambda I) = 0
$$

$$
\Rightarrow \begin{vmatrix}
R_{1,1} - \lambda & R_{1,2} \\
R_{2,1} & R_{2,2} - \lambda
\end{vmatrix} = 0
$$

$$
\Rightarrow \lambda^2 - \lambda (R_{1,1} + R_{2,2}) + (R_{1,1} R_{2,2} - R_{1,2} R_{2,1}) = 0
$$

$$
\Rightarrow \lambda_{1,2} = \frac{1}{2} \left[ (R_{1,1} + R_{2,2}) \pm \sqrt{4 R_{1,2} R_{2,1} + (R_{1,1} - R_{2,2})^2} \right]
$$
Back substitution yields the eigenvectors:

\[
\begin{bmatrix}
R_{1,1} - \lambda & R_{1,2} \\
R_{2,1} & R_{2,2} - \lambda
\end{bmatrix}
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

In general, this yields a set of linear equations. In the \( M = 2 \) case:

\[
\begin{align*}
(R_{1,1} - \lambda)q_1 + R_{1,2}q_2 &= 0 \\
R_{2,1}q_1 + (R_{2,2} - \lambda)q_2 &= 0
\end{align*}
\]

Solving the set of linear equations for a specific eigenvalue \( \lambda_i \) yields the corresponding eigenvector, \( q_i \).
Example (Two–dimensional white noise)

Let $\mathbf{R}$ be the correlation matrix of a two–sample vector of zero mean white noise

$$
\mathbf{R} = \begin{bmatrix}
\sigma^2 & 0 \\
0 & \sigma^2
\end{bmatrix}
$$

Determine the eigenvalues and eigenvectors.

Carrying out the analysis yields eigenvalues

$$
\lambda_{1,2} = \frac{1}{2} \left[ (R_{1,1} + R_{2,2}) \pm \sqrt{4R_{1,2}R_{2,1} + (R_{1,1} - R_{2,2})^2} \right] = \sigma^2
$$

and eigenvectors

$$
\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
$$

Note: The eigenvectors are unit length (and orthogonal)
Eigen Properties

Property (eigenvalues of $\mathbf{R}^k$)

If $\lambda_1, \lambda_2, \cdots, \lambda_M$ are the eigenvalues of $\mathbf{R}$, then $\lambda_1^k, \lambda_2^k, \cdots, \lambda_M^k$ are the eigenvalues of $\mathbf{R}^k$.

Proof: Note $\mathbf{R}q_i = \lambda_i q_i$. Multiplying both sides by $\mathbf{R}^{k-1}$ times,

$$\mathbf{R}^k q_i = \lambda_i \mathbf{R}^{k-1} q_i = \lambda_i^k q_i$$

Property (linear independence of eigenvectors)

The eigenvectors $q_1, q_2, \cdots, q_M$, of $\mathbf{R}$ are linearly independent, i.e.,

$$\sum_{i=1}^{M} a_i q_i \neq 0$$

for all nonzero scalars $a_1, a_2, \cdots, a_M$. 
Property (Correlation matrix eigenvalues are real & nonnegative)

The eigenvalues of $R$ are real and nonnegative.

Proof:

$$R q_i = \lambda_i q_i$$

$$\Rightarrow q_i^H R q_i = \lambda_i q_i^H q_i$$ [pre–multiply by $q_i^H$]

$$\Rightarrow \lambda_i = \frac{q_i^H R q_i}{q_i^H q_i} \geq 0$$

Follows from the facts: $R$ is positive semi-definite and $q_i^H q_i = |q_i|^2 > 0$

Note: In most cases, $R$ is positive definite and

$$\lambda_i > 0, \quad i = 1, 2, \ldots, M$$
Property (Unique eigenvalues $\Rightarrow$ orthogonal eigenvectors)

If $\lambda_1, \lambda_2, \cdots, \lambda_M$ are unique eigenvalues of $R$, then the corresponding eigenvectors, $q_1, q_2, \cdots, q_M$, are orthogonal.

Proof:

$$Rq_i = \lambda_i q_i$$

$$\Rightarrow q_j^H Rq_i = \lambda_i q_j^H q_i \quad (\ast)$$

Also, since $\lambda_j$ is real and $R$ is Hermitian

$$Rq_j = \lambda_j q_j$$

$$\Rightarrow q_j^H R = \lambda_j q_j^H$$

$$\Rightarrow q_j^H Rq_i = \lambda_j q_j^H q_i$$

Substituting the LHS from $(\ast)$

$$\Rightarrow \lambda_i q_j^H q_i = \lambda_j q_j^H q_i$$
Thus

\[ \lambda_i q_j^H q_i = \lambda_j q_j^H q_i \]
\[ \Rightarrow (\lambda_i - \lambda_j) q_j^H q_i = 0 \]

Since \( \lambda_1, \lambda_2, \ldots, \lambda_M \) are unique

\[ q_j^H q_i = 0 \quad i \neq j \]

\[ \Rightarrow q_1, q_2, \ldots, q_M \text{ are orthogonal.} \]

QED
**Diagonalization of R**

**Objective:** Find a transformation that transforms the correlation matrix into a diagonal matrix.

Let $\lambda_1, \lambda_2, \cdots, \lambda_M$ be unique eigenvectors of $R$ and take $q_1, q_2, \cdots, q_M$ to be the $M$ orthonormal eigenvectors

$$q_i^H q_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Define $Q = [q_1, q_2, \cdots, q_M]$ and $\Omega = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_M)$. Then consider

$$Q^H R Q = \begin{bmatrix} q_1^H \\ q_2^H \\ \vdots \\ q_M^H \end{bmatrix} R [q_1, q_2, \cdots, q_M]$$
\[ Q^H R Q = \begin{bmatrix} q_1^H & q_2^H & \cdots & q_M^H \end{bmatrix} R [q_1, q_2, \cdots, q_M] \]

\[ = \begin{bmatrix} q_1^H & q_2^H & \cdots & q_M^H \end{bmatrix} [\lambda_1 q_1, \lambda_2 q_2, \cdots, \lambda_N q_M] \]

\[ = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M \end{bmatrix} \]

\[ \Rightarrow Q^H R Q = \Omega \quad \text{(eigenvector diagonalization of } R) \]
Property (Q is unitary)

Q is unitary, i.e., $Q^{-1} = Q^H$

Proof: Since the $q_i$ eigenvectors are orthonormal

$$Q^H Q = \begin{bmatrix} q_1^H & q_2^H & \cdots & q_M^H \end{bmatrix} [q_1, q_2, \cdots, q_M] = I$$

$\Rightarrow Q^{-1} = Q^H$

Property (Eigen decomposition of R)

The correlation matrix can be expressed as

$$R = \sum_{i=1}^{M} \lambda_i q_i q_i^H$$
Proof: The correlation diagonalization result states

\[ Q^H R Q = \Omega \]

Isolating \( R \) and expanding,

\[ R = Q \Omega Q^H = [q_1, q_2, \ldots, q_M] \Omega \]

\[ = [q_1, q_2, \ldots, q_M] \begin{bmatrix} \lambda_1 q_1^H \\ \lambda_2 q_2^H \\ \vdots \\ \lambda_M q_M^H \end{bmatrix} = \sum_{i=1}^{M} \lambda_i q_i q_i^H \]

Note: This also gives

\[ R^{-1} = (Q^H)^{-1} \Omega^{-1} Q^{-1} = Q \Omega^{-1} Q^H \]

where \( \Omega^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_M) \)
Aside (trace & determinant for matrix products)

Note $\text{trace}(A) \equiv \sum_i A_{i,i}$. Also,

$$\text{trace}(AB) = \text{trace}(BA) \quad \text{similarly} \quad \text{det}(AB) = \text{det}(A)\text{det}(B)$$

Property (Determinant–Eigenvalue Relation)

The determinant of the correlation matrix is related to the eigenvalues as follows:

$$\text{det}(R) = \prod_{i=1}^{M} \lambda_i$$

Proof: Using $R = \mathbf{Q}\Omega\mathbf{Q}^H$ and the above,

$$\text{det}(R) = \text{det}(\mathbf{Q}\Omega\mathbf{Q}^H) = \text{det}(\mathbf{Q})\text{det}(\mathbf{Q}^H)\text{det}(\Omega) = \text{det}(\Omega) = \prod_{i=1}^{M} \lambda_i$$
Property (Trace–Eigenvalue Relation)

The trace of the correlation matrix is related to the eigenvalues as follows:

\[
\text{trace}(R) = \sum_{i=1}^{M} \lambda_i
\]

Proof: Note

\[
\text{trace}(R) = \text{trace}(Q \Omega Q^H) = \text{trace}(Q^H \Omega Q) = \text{trace}(\Omega) = \sum_{i=1}^{M} \lambda_i
\]

QED
Definition (Normal Matrix)
A complex square matrix \( A \) is a normal matrix if

\[
A^H A = AA^H
\]

That is, a matrix is normal if it commutes with its conjugate transpose.

Note
- All Hermitian symmetric matrices are normal
- Every matrix that can be diagonalized by the unitary transform is normal

Definition (Condition Number)
The condition number reflects how numerically well–conditioned a problem is, i.e, a low condition number ⇒ well–conditioned; a high condition number ⇒ ill–conditioned.
Definition (Condition Number for Linear Systems)

For a linear system

\[ A\mathbf{x} = \mathbf{b} \]

defined by a normal matrix \( A \), the condition number is

\[ \chi(A) = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \]

where \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) are the maximum/minimum eigenvalues of \( A \)

Observations:

- Large eigenvalue spread \( \Rightarrow \) ill–conditioned
- Small eigenvalue spread \( \Rightarrow \) well–conditioned
The discrete Karhmen-Loeve Transform (KLT)

Definition (The discrete Karhmen-Loeve Transform (KLT))

A $M$ sample vector $x(n)$ from the process $\{x(n)\}$ can be expressed as

$$x(n) = \sum_{i=1}^{M} c_i(n) q_i$$

where $q_1, q_2, \cdots, q_M$ are the orthonormal eigenvectors of the process correlation matrix, $R$, and $c_1(n), c_2(n), \cdots, c_M(n)$ are a set of KLT coefficients.

- Signal is represented as a weighted sum of eigenvectors
- Need to determine the coefficients
Determining Coefficients: Write the expression in matrix form

\[ x(n) = \sum_{i=1}^{M} c_i(n) q_i \]

\[ = Qc(n) \quad (\star) \]

where

\[ Q = [q_1, q_2, \cdots, q_M] \]

and

\[ c(n) = [c_1(n), c_2(n), \cdots, c_M(n)]^T. \]

Solving (\star) for \( c(n) \):

\[ c(n) = Q^{-1}x(n) = Q^Hx(n) \]

or

\[ c_i(n) = q_i^Hx(n) \]

Note: \( c_i(n) \) is the projection of \( x(n) \) onto \( q_i \)
Question: How related are the coefficients to reach other?

Answer: Consider the correlation between \( c_i(n) \) terms

\[
R_{c(n)} = E\{c(n)c^H(n)\} = E\{(Q^Hx(n))(Q^Hx(n))^H\} = E\{(Q^Hx(n)x^H(n)Q\} = Q^H R_x Q = \Omega
\]

Result:

\[
E\{c_i^*(n)c_j(n)\} = \begin{cases} \lambda_i & i = j \\ 0 & \text{otherwise} \end{cases}
\]

\( \Rightarrow \) KLT transform coefficients are uncorrelated

- A desirable property – Why?
**Question:** Can we represent $x(n)$ with fewer terms? If so, how do we minimize the representation error?

**Approach:** Use fewer terms in the KLT transform

$$x(n) = \sum_{i=1}^{M} c_i(n) q_i$$

$$\Rightarrow \hat{x}(n) = \sum_{i=1}^{N} c_i(n) q_i \quad N < M$$

Thus

$$x(n) = \hat{x}(n) + \epsilon(n)$$

$$= \sum_{i=1}^{N} c_i(n) q_i + \sum_{i=N+1}^{M} c_i(n) q_i$$

**Question:** How do we minimize the representation error?
**Approach:** Analyzed and minimize the error power

The error power is given by

\[
\varepsilon = E\{\varepsilon^H(n)\varepsilon(n)\} = E\left\{ \sum_{i=N+1}^{M} c_i^*(n)q_i^H \sum_{j=N+1}^{M} c_j(n)q_j \right\} 
\]

\[
= \sum_{i=N+1}^{M} E\{c_i^*(n)c_i(n)\} \quad [\text{result of orthogonality}]
\]

\[
= \sum_{i=N+1}^{M} \lambda_i \quad [\text{from prior result}]
\]

**Result:** To minimize the error select the \(q_i\) eigenvectors associated with \(M\) largest eigenvalues.