ELEG-636: Statistical Signal Processing

Gonzalo R. Arce

Department of Electrical and Computer Engineering
University of Delaware

Spring 2010
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: $R$ 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 $R$, find a vector $q$ satisfying

$$Rq = \lambda q$$

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

$$Rq_i = \lambda_i q_i \quad i = 1, 2, 3, \ldots, M$$

To see this, note

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

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

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

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

\[
Rq_i = \lambda_i q_i
\]

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

\[
R(q_i) = \lambda_i q_i
\]

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

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

Let \( M = 2 \) and

\[
\mathbf{R} = \begin{bmatrix}
R_{1,1} & R_{1,2} \\
R_{2,1} & R_{2,2}
\end{bmatrix}
\]

Determine the eigenvalues and eigenvectors.

Thus

\[
det(\mathbf{R} - \lambda \mathbf{I}) = 0
\]

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

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

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

\[
(R_{1,1} - \lambda)q_1 + R_{1,2}q_2 = 0
\]
\[
R_{2,1}q_1 + (R_{2,2} - \lambda)q_2 = 0
\]

- 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

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

\[
= \frac{1}{2} \left[ (\sigma^2 + \sigma^2) \pm \sqrt{0 + (\sigma^2 - \sigma^2)} \right] = \sigma^2
\]

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 $R^k$)

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

Proof: Note $Rq_i = \lambda_i q_i$. Multiplying both sides by $R$ $k-1$ times,

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

Property (linear independence of eigenvectors)

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

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

for all nonzero scalars $a_1, a_2, \cdots, a_M$. 
Eigen Properties

Property (Correlation matrix eigenvalues are real & nonnegative)

The eigenvalues of $\mathbf{R}$ are real and nonnegative.

Proof:

$$\mathbf{Rq}_i = \lambda_i \mathbf{q}_i$$

$\Rightarrow \mathbf{q}_i^H \mathbf{Rq}_i = \lambda_i |\mathbf{q}_i|^2$ [pre-multiply by $\mathbf{q}_i^H$]

$\Rightarrow \lambda_i \geq 0$

Follows from the facts: $\mathbf{R}$ is positive semi-definite and $\mathbf{q}_i^H \mathbf{q}_i = |\mathbf{q}_i|^2 > 0$

Note: In most cases, $\mathbf{R}$ is positive definite and

$$\lambda_i > 0, \quad i = 1, 2, \cdots, M$$
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:

\[
Rq_i = \lambda_i q_i \\
\Rightarrow q_j^H R q_i = \lambda_i q_j^H q_i \tag{*}
\]

Also, since \(\lambda_j\) is real and \(R\) is Hermitian

\[
R q_j = \lambda_j q_j \\
\Rightarrow q_j^H R = \lambda_j q_j^H \\
\Rightarrow q_j^H R q_i = \lambda_j q_j^H q_i
\]

Substituting the LHS from (*)

\[
\Rightarrow \lambda_i q_j^H q_i = \lambda_j q_j^H q_i
\]
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 \( \lambda_1, \lambda_2, \cdots, \lambda_M \) are unique

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

\[ \Rightarrow q_1, q_2, \cdots, q_M \] are orthogonal.

QED
Diagonalization of $\mathbf{R}$

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

Let $\lambda_1, \lambda_2, \cdots, \lambda_M$ be unique eigenvalues of $\mathbf{R}$ and take $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$ to be the $M$ orthonormal eigenvectors

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

Define $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M]$ and $\Omega = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_M)$. Then consider

$$\mathbf{Q}^H \mathbf{R} \mathbf{Q} = \begin{bmatrix} q_1^H \\ q_2^H \\ \vdots \\ q_M^H \end{bmatrix} \begin{bmatrix} \mathbf{R}[\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M] \end{bmatrix}$$
\[ Q^H R Q = \begin{bmatrix} q_1^H \\ q_2^H \\ \vdots \\ q_M^H \end{bmatrix} R [q_1, q_2, \ldots, q_M] \]

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

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

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

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

**Proof:** Since the $q_i$ eigenvectors are orthonormal

\[
Q^H Q = \begin{bmatrix}
q_1^H \\
q_2^H \\
\vdots \\
q_M^H
\end{bmatrix}
[q_1, q_2, \ldots, q_M] = I
\]

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

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

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

Property (Determinant–Eigenvalue Relation)

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

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

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

$$\det(R) = \det(Q\Omega Q^H)$$

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

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

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

Proof: Note

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

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 $\Rightarrow$ well–conditioned; a high condition number $\Rightarrow$ ill–conditioned.
Definition (Condition Number for Linear Systems)

For a linear system

\[ Ax = b \]

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
**Sensitivity Analysis**: Suppose a system (filter) is related as:

\[ R w = d \]

where \( w \) defines the filter parameters; \( R \) and \( d \) are signal statistic matrices

⇒ Introduce small signal statistic perturbations

Let \( R \) and \( d \) are perturbed such that \( \| \delta R \| / \| R \| \) and \( \| \delta d \| / \| d \| \) are on the order \( \epsilon \ll 1 \)

**Result**: A bound on the resulting parameter perturbations is given by

\[ \frac{\| \delta w \|}{\| w \|} \leq \epsilon \chi(R) = \epsilon \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \]

**Consequence**: If \( R \) is ill conditioned, small changes in \( R \) or \( d \) can lead to big changes in \( w \)

⇒ System sensitivity is related to eigenvalue spread
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

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

\[ = Qc(n) \quad \text{(*)} \]

where

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

and

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

Solving (*) 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

\[
\begin{align*}
R_c(n) &= E\{c(n)c^H(n)\} \\
&= E\{(Q^Hx(n))(Q^Hx(n))^H\} \\
&= E\{(Q^Hx(n)x^H(n)Q\} \\
&= Q^H R_x Q \\
&= \Omega
\end{align*}
\]

**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

\[
\epsilon = \mathbb{E}\{\epsilon^H(n)\epsilon(n)\} \\
= \mathbb{E}\left\{ \sum_{i=N+1}^{M} c_i^*(n) q_i^H \sum_{j=N+1}^{M} c_j(n) q_j \right\} \\
= \sum_{i=N+1}^{M} \mathbb{E}\{c_i^*(n)c_i(n)\} \quad \text{[result of orthogonality]} \\
= \sum_{i=N+1}^{M} \lambda_i \quad \text{[from prior result]}
\]

**Result:** To minimize the error select the \( q_i \) eigenvectors associated with \( M \) largest eigenvalues.
**Objective:** Find the optimal filter coefficients for two cases: (1) deterministic signals and (2) stochastic signals

**Signal model:**

\[ x(n) = u(n) + v(n) \]

- \( u(n) \): signal
- \( v(n) \): noise
Definitions:

Filter parameters: \( \mathbf{w} = [w_1, w_2, \cdots, w_M]^T \)

Observation vector: \( \mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-M+1)]^T \)

Then using \( x(n) = u(n) + v(n) \) and linearity,

\[
\begin{align*}
y(n) &= \mathbf{w}^T \mathbf{x}(n) \\
&= \mathbf{w}^T \mathbf{u}(n) + \mathbf{w}^T \mathbf{v}(n) \\
&= y_s(n) + y_n(n)
\end{align*}
\]

where

\[
\begin{align*}
y_s(n) &= \mathbf{w}^T \mathbf{u}(n) \\
y_n(n) &= \mathbf{w}^T \mathbf{v}(n)
\end{align*}
\]

- Consider deterministic and stochastic cases separately
Case 1: Matched Filters for Deterministic Signals

Approach:

Analyzed output SNR
Set filter coefficients to maximize SNR

The SNR at time $n$ can be defined as

$$\text{SNR} = \frac{|y_s(n)|^2}{E\{|y_n(n)|^2\}}$$

where

$$y_s(n) = w^T u(n) \quad \text{[Deterministic]}$$
$$y_n(n) = w^T v(n) \quad \text{[Stochastic]}$$

Note: Case assumes $u(n)$ is deterministic
Using \( y_s(n) = w^T u(n) \) and \( y_n(n) = w^T v(n) \),

\[
\text{SNR} = \frac{|y_s(n)|^2}{E\{|y_n(n)|^2\}} = \frac{(w^T u(n))^* (u^T (n) w)}{E\{(w^T v(n))^* (v^T (n) w)\}} = \frac{w^H u^* (n) u^T (n) w}{E\{w^H v^* (n) v^T (n) w\}} = \frac{w^H u^* (n) u^T (n) w}{w^H R_v w}
\]

- To analyze further, restrict the noise statistics.
Suppose $v(n)$ is zero mean and uncorrelated, $R_v = \sigma_v^2 I$. Then

$$\text{SNR} = \frac{w^H u^*(n) u^T(n)w}{w^H R_v w} = \frac{|w^H u^*(n)|^2}{\sigma_v^2 w^H w}$$

If we restrict $|w|^2 = 1$, then the final result is

$$\text{SNR} = \frac{|w^H u^*(n)|^2}{\sigma_v^2}$$
Definition (Cauchy–Schwarz Inequality (1821 disc.; 1859 cont.))

Cauchy-Schwarz’s inequality states (matrix case)

\[ |A^H B|^2 \leq (A^H A)(B^H B) \]

with equality only if \( A = kB \), where \( k \) is a constant

Thus

\[
\text{SNR} = \frac{|w^H u^*(n)|^2}{\sigma_v^2} \leq \frac{(w^H w)(u^T(n)u^*(n))}{\sigma_v^2}
\]

\[
= \frac{u^H(n)u(n)}{\sigma_v^2} = \frac{|u(n)|^2}{\sigma_v^2}
\]

Result: The SNR is maximized when \( w = ku^*(n) \)
Example

A communications system sends out one of two signals

Symbol “1”

Symbol “0”

We receive

\[ y(n) = u_i(n) + v(n) \]

where \( v(n) \) is i.i.d. Gaussian.

**Objective:** Find the optimal \( w \) (assume \( n = 3 \))
Set $w = k u_1(n)$ where $k$ is such that $|w|^2 = 1$. Thus,

$$w = u_1(n)/|u_1(n)| = \frac{1}{\sqrt{10}}[2, 2, 1, 1]^T$$

The resulting impulse response of the filter is

$$y_s(n) = w^T u_1(n) = \frac{|u_1(n)|^2}{|u_1(n)|} = \sqrt{10}$$

**Observation:** Optimal filter impulse response is sent (desired) signal time–reversed & normalized

Determine $y_s(n)$ for signal “1” sent
For signal “0”

\[ y_s(n) = w^T u_0(n) = \frac{-|u_0(n)|^2}{|u_1(n)|} = -\sqrt{10} \]

Received signal distributions:

The SNR for the optimized system is

\[ \text{SNR} = \frac{|u_1(n)|^2}{\sigma_v^2} = \frac{10}{\sigma_v^2} \]
Case 2: Matched Filter for Stochastic Signals

As before

\[ y(n) = w^T x(n) = w^T u(n) + w^T v(n) = y_s(n) + y_n(n) \]

Now define the SNR as

\[ \text{SNR} = \frac{E\{|y_s(n)|^2\}}{E\{|y_n(n)|^2\}} = \frac{w^H R_u w}{w^H R_v w} \]

As before, if \( R_v = \sigma_v^2 I \) and \( |w|^2 = 1 \)

\[ \text{SNR} = \frac{w^H R_u w}{\sigma_v^2} \]
Since

\[ \text{SNR} = \frac{w^H R_u w}{\sigma_v^2}, \]

maximizing the SNR is equivalent to maximizing \( |w^H R_u w| \)

Recall from Cauchy–Schwarz,

\[ |A^H B|^2 \leq (A^H A)(B^H B) \]

with equality only if \( A = kB \)

Let \( A^H = w^H R_u \) and \( B = w \) and apply Cauchy–Schwarz

Result: \( |w^H R_u w| \) is maximized for \( A = kB \)

\[ \Rightarrow R_u w = kw \]

\[ \Rightarrow k \text{ is an eigenvalue and } w \text{ is an eigenvector!} \]

\[ \bullet \ w = q_i \text{ and } k = \lambda_i, \text{ where } R_u q_i = \lambda_i q_i \]

Question: How do we choose which eigenvector?
Using \( w = q_i \), or \( w^H = q_i^H \), and \( R_u w = R_u q_i = \lambda_i q_i \) gives

\[
\text{SNR} = \frac{w^H R_u w}{\sigma_v^2} = \frac{q_i^H \lambda_i q_i}{\sigma_v^2} \quad \text{[orthonormal eigenvectors]}
\]

\[
= \frac{\lambda_i}{\sigma_v^2}
\]

**Result:** The SNR is maximized when \( w = q_{\text{max}} \), the eigenvector associated with the largest eigenvalue of \( R_u \).

- When using the optimal filter weights \( (w = q_{\text{max}}) \), the resulting (maximized) SNR is

\[
\text{SNR} = \frac{\lambda_{\text{max}}}{\sigma_v^2}
\]