FSAN-815/ELEG-815: Foundations of Statistical Learning

Gonzalo R. Arce

Chapter 3: Eigen Analysis

Fall 2014

Course Objectives & Structure

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.

Eigen Analysis

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

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

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: $\det(\mathbf{R} - \lambda\mathbf{I})$ is a Mth order polynomial in λ

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

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

- Each eigenvector $\mathbf{q}_i$ is associated with one eigenvalue λ_i
- The eigenvectors are not unique

$$\begin{aligned} \mathbf{R}\mathbf{q}_i &= \lambda_i \mathbf{q}_i \\ \Rightarrow \mathbf{R}(a\mathbf{q}_i) &= \lambda_i (a\mathbf{q}_i) \end{aligned}$$

Consequence: eigenvectors are generally normalized, e.g., $|\mathbf{q}_i| = 1$ 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

$$\begin{aligned} \det(\mathbf{R} - \lambda \mathbf{I}) &= 0 \\ \Rightarrow \begin{vmatrix} R_{1,1} - \lambda & R_{1,2} \\ R_{2,1} & R_{2,2} - \lambda \end{vmatrix} &= 0 \\ \Rightarrow \lambda^2 - \lambda(R_{1,1} + R_{2,2}) + (R_{1,1}R_{2,2} - R_{1,2}R_{2,1}) &= 0 \end{aligned}$$

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

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:

$$\begin{aligned} (R_{1,1}-\lambda)q_1 + R_{1,2}q_2 &= 0 \\ R_{2,1}q_1 + (R_{2,2}-\lambda)q_2 &= 0 \end{aligned}$$

- Solving the set of linear equations for a specific eigenvalue λ_i yields the corresponding eigenvector, $\mathbf{q}_i$

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

$$\begin{aligned} \lambda_{1,2} &= \frac{1}{2}\left[(R_{1,1}+R_{2,2}) \pm \sqrt{4R_{1,2}R_{2,1}+(R_{1,1}-R_{2,2})}\right] \\ &= \frac{1}{2}\left[(\sigma^2+\sigma^2) \pm \sqrt{0+(\sigma^2-\sigma^2)}\right] = \sigma^2 \end{aligned}$$

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 $\mathbf{R}\mathbf{q}_i = \lambda_i \mathbf{q}_i$.Multiplying both sides by $\mathbf{R}$ $k-1$ times,

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

Property (linear independence of eigenvectors)

The eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$, of $\mathbf{R}$ are linearly independent, i.e.,

$$\sum_{i=1}^{M} a_i \mathbf{q}_i \neq \mathbf{0}$$

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

Property (Correlation matrix eigenvalues are real & nonnegative)

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

Proof:

$$\begin{aligned} \mathbf{R}\mathbf{q}_i &= \lambda_i \mathbf{q}_i \\ \Rightarrow \mathbf{q}_i^H \mathbf{R}\mathbf{q}_i &= \lambda_i \mathbf{q}_i^H \mathbf{q}_i \qquad \text{[pre–multiply by } \mathbf{q}_i^H\text{]} \\ \Rightarrow \lambda_i &= \frac{\mathbf{q}_i^H \mathbf{R}\mathbf{q}_i}{\mathbf{q}_i^H \mathbf{q}_i} \geq 0 \end{aligned}$$

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

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

Property (Unique eigenvalues $\Rightarrow$ orthogonal eigenvectors)

If $\lambda_1, \lambda_2, \cdots, \lambda_M$ are unique eigenvalues of $\mathbf{R}$, then the corresponding eigenvectors, $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$, are orthogonal.

Proof:

$$\begin{aligned} \mathbf{R}\mathbf{q}_i &= \lambda_i \mathbf{q}_i \\ \Rightarrow \mathbf{q}_j^H \mathbf{R}\mathbf{q}_i &= \lambda_i \mathbf{q}_j^H \mathbf{q}_i \qquad (*) \end{aligned}$$

Also, since λ_j is real and $\mathbf{R}$ is Hermitian

$$\begin{aligned} \mathbf{R}\mathbf{q}_j &= \lambda_j \mathbf{q}_j \\ \Rightarrow \mathbf{q}_j^H \mathbf{R} &= \lambda_j \mathbf{q}_j^H \\ \Rightarrow \mathbf{q}_j^H \mathbf{R}\mathbf{q}_i &= \lambda_j \mathbf{q}_j^H \mathbf{q}_i \end{aligned}$$

Substituting the LHS from $(*)$

$$\Rightarrow \lambda_i \mathbf{q}_j^H \mathbf{q}_i = \lambda_j \mathbf{q}_j^H \mathbf{q}_i$$

Thus

$$\begin{aligned} \lambda_i \mathbf{q}_j^H \mathbf{q}_i &= \lambda_j \mathbf{q}_j^H \mathbf{q}_i \\ \Rightarrow (\lambda_i - \lambda_j) \mathbf{q}_j^H \mathbf{q}_i &= 0 \end{aligned}$$

Since $\lambda_1, \lambda_2, \cdots, \lambda_M$ are unique

$$\mathbf{q}_j^H \mathbf{q}_i = 0 \qquad i \neq j$$

$\Rightarrow \mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$ 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 $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$ to be the M orthonormal eigenvectors

$$\mathbf{q}_i^H \mathbf{q}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

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

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

$$
\begin{aligned}
\mathbf{Q}^H\mathbf{R}\mathbf{Q} &= \begin{bmatrix} \mathbf{q}_1^H \\ \mathbf{q}_2^H \\ \vdots \\ \mathbf{q}_M^H \end{bmatrix} \mathbf{R}[\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M] \\
&= \begin{bmatrix} \mathbf{q}_1^H \\ \mathbf{q}_2^H \\ \vdots \\ \mathbf{q}_M^H \end{bmatrix} [\lambda_1\mathbf{q}_1, \lambda_2\mathbf{q}_2, \cdots, \lambda_N\mathbf{q}_M] \\
&= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M \end{bmatrix} \\
\Rightarrow \mathbf{Q}^H\mathbf{R}\mathbf{Q} &= \mathbf{\Omega} \quad \text{(eigenvector diagonalization of } \mathbf{R}\text{)}
\end{aligned}
$$

Property ($\mathbf{Q}$ is unitary)

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

Proof: Since the $\mathbf{q}_i$ eigenvectors are orthonormal

$$\begin{aligned} \mathbf{Q}^H\mathbf{Q} &= \begin{bmatrix} \mathbf{q}_1^H \\ \mathbf{q}_2^H \\ \vdots \\ \mathbf{q}_M^H \end{bmatrix} [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M] = \mathbf{I} \\ \Rightarrow \mathbf{Q}^{-1} &= \mathbf{Q}^H \end{aligned}$$

Property (Eigen decomposition of $\mathbf{R}$)

The correlation matrix can be expressed as

$$\mathbf{R} = \sum_{i=1}^{M} \lambda_i \mathbf{q}_i \mathbf{q}_i^H$$

Proof: The correlation diagonalization result states

$$\mathbf{Q}^H \mathbf{R} \mathbf{Q} = \mathbf{\Omega}$$

Isolating **R** and expanding,

$$\begin{aligned}
\mathbf{R} &= \mathbf{Q}\mathbf{\Omega} Q^H = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M]\mathbf{\Omega} \begin{bmatrix} \mathbf{q}_1^H \\ \mathbf{q}_2^H \\ \vdots \\ \mathbf{q}_M^H \end{bmatrix} \\
&= [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M] \begin{bmatrix} \lambda_1 \mathbf{q}_1^H \\ \lambda_2 \mathbf{q}_2^H \\ \vdots \\ \lambda_M \mathbf{q}_M^H \end{bmatrix} = \sum_{i=1}^{M} \lambda_i \mathbf{q}_i \mathbf{q}_i^H
\end{aligned}$$

Note: This also gives

$$\mathbf{R}^{-1} = (\mathbf{Q}^H)^{-1} \mathbf{\Omega}^{-1} \mathbf{Q}^{-1} = \mathbf{Q}\mathbf{\Omega}^{-1}\mathbf{Q}^H$$

where $\mathbf{\Omega}^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \cdots, 1/\lambda_M)$

Aside (trace & determinant for matrix products)

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

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

Property (Determinant–Eigenvalue Relation)

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

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

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

$$\begin{aligned} \det(\mathbf{R}) &= \det(\boldsymbol{Q\Omega Q}^H) \\ &= \det(\mathbf{Q})\det(\mathbf{Q}^H)\det(\boldsymbol{\Omega}) = \det(\boldsymbol{\Omega}) = \prod_{i=1}^{M} \lambda_i \end{aligned}$$

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

$$\begin{aligned}
\text{trace}(\mathbf{R}) &= \text{trace}(\boldsymbol{Q}\boldsymbol{\Omega}\boldsymbol{Q}^H) \\
&= \text{trace}(\mathbf{Q}^H\boldsymbol{Q}\boldsymbol{\Omega}) \\
&= \text{trace}(\boldsymbol{\Omega}) \\
&= \sum_{i=1}^{M} \lambda_i
\end{aligned}$$

QED

Definition (Normal Matrix)

A complex square matrix $\mathbf{A}$ is a normal matrix if

$$\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$$

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

$$\mathbf{Ax} = \mathbf{b}$$

defined by a normal matrix **A**, the condition number is

$$\chi(\mathbf{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ 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 $\mathbf{x}(n)$ from the process $\{x(n)\}$ can be expressed as

$$\mathbf{x}(n) = \sum_{i=1}^{M} c_i(n)\mathbf{q}_i$$

where $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M$ are the orthonormal eigenvectors of the process correlation matrix, $\mathbf{R}$, 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

$$\begin{aligned}\mathbf{x}(n) &= \sum_{i=1}^{M} c_i(n)\mathbf{q}_i \\ &= \mathbf{Q}\mathbf{c}(n) \qquad\qquad (*)\end{aligned}$$

where

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M]$$

and

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

Solving $(*)$ for $\mathbf{c}(n)$:

$$\mathbf{c}(n) = \mathbf{Q}^{-1}\mathbf{x}(n) = \mathbf{Q}^H\mathbf{x}(n)$$

or

$$c_i(n) = \mathbf{q}_i^H\mathbf{x}(n)$$

Note: $c_i(n)$ is the projection of $\mathbf{x}(n)$ onto $\mathbf{q}_i$

Question: How related are the coefficients to reach other?

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

$$\begin{aligned}
\mathbf{R}_{\mathbf{c}(n)} &= E\{\mathbf{c}(n)\mathbf{c}^H(n)\} \\
&= E\{(\mathbf{Q}^H\mathbf{x}(n))(\mathbf{Q}^H\mathbf{x}(n))^H\} \\
&= E\{(\mathbf{Q}^H\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{Q}\} \\
&= \mathbf{Q}^H\mathbf{R}_x\mathbf{Q} \\
&= \mathbf{\Omega}
\end{aligned}$$

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 $\mathbf{x}(n)$ with fewer terms?If so, how do we minimize the representation error?

Approach: Use fewer terms in the KLT transform

$$\begin{aligned} \mathbf{x}(n) &= \sum_{i=1}^{M} c_i(n)\mathbf{q}_i \\ \Rightarrow \hat{\mathbf{x}}(n) &= \sum_{i=1}^{N} c_i(n)\mathbf{q}_i \quad N < M \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{x}(n) &= \hat{\mathbf{x}}(n) + \varepsilon(n) \\ &= \sum_{i=1}^{N} c_i(n)\mathbf{q}_i + \sum_{i=N+1}^{M} c_i(n)\mathbf{q}_i \end{aligned}$$

Question: How do we minimize the representation error?

Approach: Analyzed and minimize the error power

The error power is given by

$$\begin{aligned} \varepsilon &= E\{\varepsilon^H(n)\varepsilon(n)\} \\ &= E\left\{ \sum_{i=N+1}^{M} c_i^*(n)\mathbf{q}_i^H \sum_{j=N+1}^{M} c_j(n)\mathbf{q}_j \right\} \\ &= \sum_{i=N+1}^{M} E\{c_i^*(n)c_i(n)\} \qquad \text{[result of orthogonality]} \\ &= \sum_{i=N+1}^{M} \lambda_i \qquad \text{[from prior result]} \end{aligned}$$

Result: To minimize the error select the $\mathbf{q}_i$ eigenvectors associated with M largest eigenvalues.