MODULE 1.1
Spectral decomposition of a real symmetric matrix

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

Eigenvalue and eigenvector pair of a matrix

- Let $A \in R^{n \times n}$ be a real matrix of order n
- If there exist a scalar, $\lambda(real/complex)$ and a vector, v(real/complex) such that

$$Av = \lambda v \tag{1}$$

  then λ is the eigenvalue and v is the corresponding eigenvector of A
- The pair (λ, v) satisfying (1) is called an eigenpair of A
- The set of all eigenvalues of A is called the spectrum of A

Invariant subspace of A

- Let $S_k = \{v_1, v_2, ..., v_k\}$ be a set of linearly independent vectors in R^n
- $SPAN(S_k)$ denotes the set of all linear combinations of the vectors in S_k
- $SPAN(S_k)$ is a K-dimensional subspace of R^n
- If $AX \in SPAN(S_k)$ for any $X \in SPAN(S_k)$, then S_k is said to be A-invariant
- From (1), since $Av \in SPAN(v)$, every eigenvector defines an invariant subspace of dimension 1.

- Rewrite(1) as a linear homogeneous system:

$$(A - \lambda I)v = 0 \tag{2}$$

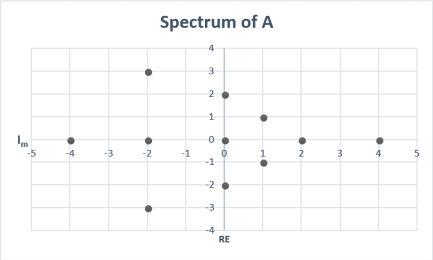
- Equation (2) has a non-null solution, exactly when $(A - \lambda I)$ is singular, that is

$$p(\lambda) = |A - \lambda I| = 0 \tag{3}$$

- The n eigenvalues of A are given by the n roots of the characteristic polynomial, $p(\lambda)$ of A

Distribution of eigenvalues of A

- Since A is real, the coefficients of $p(\lambda)$ are also real
- An n^{th} degree polynomial of degree n has n roots
- The roots are real or complex and the complex roots occur in conjugate pairs
- A typical distribution of eigenvalues



Eigenpairs of a real symmetric matrix

- Let $A \in R^{n \times n}$ and $A^T = A$, that is, A is symmetric
- SM1: The eigenvalues of a real symmetric matrix are real
- SM2: Eigenvectors corresponding to distinct eigenvalues are orthogonal
- SM3: If λ as a root of $p(\lambda) = 0$ in (3) is of (algebraic) multiplicity $1 \leq k \leq n$, then there exists a set of k mutually orthogonal vectors $v_1, v_2, v_3, ..., v_k$ such that (λ, v_i) is an eigenpair of A for $1 \leq i \leq k$, that is, k is also the geometric multiplicity which is the dimension of the invariant subspace spanned by $\{v_1, v_2, ...v_k\}$ where $Av_i = \lambda v_i$ for $1 \leq i \leq k$

Matrix of eigenvalues and eigenvectors

- Let (λ_i, v_i) such that
$$Av_i = \lambda_i v_i \tag{4}$$
- Define
$$V = [v_1, v_2, ..., v_n] \in R^{n\times n}$$
$$\Lambda = Diag[\lambda_1, \lambda_2, ..., \lambda_n] \in R^{n\times n}$$
- Then (4) becomes:
$$AV = V\Lambda \tag{5}$$
- The eigenvectors are mutually orthogonal (see appendix)
$$v_i^T v_j \neq 0 \quad for \quad i = j$$
$$= 0 \quad \text{otherwise}$$

Orthonormality of eigenvectors

- Since $Av = \lambda v \implies A(\alpha v) = \lambda(\alpha v)$ for any α, non-zero constant, we need to only consider unit vector for eigenvectors.
- Consequently, assume that the vectors v_i in (4) are orthonormal:

$$\begin{aligned} v_i^T v_j &= 1 \quad \textit{if } i = j \\ &= 0 \quad \textit{otherwise} \end{aligned} \tag{6}$$

- Hence, V is an orthogonal matrix, that is, using (6):

$$V^TV = \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \\ \vdots \\ v_n^T \end{bmatrix} \begin{bmatrix} v_1, v_2, v_3, \ldots, v_n \end{bmatrix}$$

$$= \begin{bmatrix} v_1^T v_1 & v_1^T v_2 & \ldots & v_1^T v_n \\ v_2^T v_1 & v_2^T v_2 & \ldots & v_2^T v_n \\ v_3^T v_1 & v_3^T v_2 & \ldots & v_3^T v_n \\ & & \vdots & \\ v_n^T v_1 & v_n^T v_2 & \ldots & v_n^T v_n \end{bmatrix}$$

$$= I_n = VV^T \tag{7}$$

Spectral or eigen decomposition of a symmetric matrix

- Multiplying both sides of (5) by V^T and using (7), we obtain

$$A = AVV^T = V\Lambda V^T \tag{8}$$

- This multiplicative decomposition in (8) is called the eigen decomposition of A

- Expanding V and Λ in (8): A =

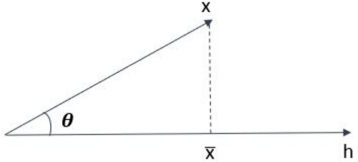
$$\begin{bmatrix} v_1, v_2, v_3, \ldots, v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i V_i V_i^T \tag{9}$$

- Since $V_i V_i^T$ is a rank-1 (outer product) matrix, (9) expresses A as a sum of n linearly independent rank-1 matrices

A digression

- Consider:



- Let $\hat{h} = \frac{h}{||h||}$ be the unit vector along h
- Orthogonal projection, $\bar{x}$ of x along h is given by

$$\bar{x} = (x^T\hat{h})\hat{h} = (\hat{h}^T x)\hat{h} = \hat{h}(\hat{h}^T x) = (\hat{h}\hat{h}^T)x \qquad (10)$$

- The rank-1 matrix

$$P_h = (\hat{h}\hat{h}^T) \qquad (11)$$

is called an orthogonal projection matrix and (10) becomes:

$$\bar{x} = P_h x$$

- Consequently, the rank-1 matrix $v_i v_i^T$ in (9) is an orthogonal projection matrix along v_i
- That is, (9) expresses A as a linear combination of orthogonal projection matrices

A-symmetric and positive definite (SPD)

- In this case, the eigenvalues of A are all real and positive
- That is, we can express

$$\Lambda = \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} \tag{12}$$

where

$$\Lambda^{\frac{1}{2}} = Diag(\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \ldots, \lambda_n^{\frac{1}{2}})$$

- $A = V\Lambda V^T = V\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}V^T$

$$= (V\Lambda^{\frac{1}{2}})(V\Lambda^{\frac{1}{2}})^T = \bar{V}\bar{V}^T \tag{13}$$

is the another form of the eigen decomposition for A

Why SPD matrices?

- In multivariate statistical analysis, SPD matrices arise naturally as covariance matrices
- In fact, the many well known methods in multivariate statistical analysis such as
  - Principal Component Analysis (PCA)
  - Singular Value Decomposition (SVD)
  - Cononical Correlation (CC)

  are based on the spectral or eigen decomposition of SPD matrices

- The goal of this appendix is to provide a proof of various properties of real symmetric matrices used in the development of this module
- The final result is to prove that every real symmetric matrix is diagonalizable using orthogonal transformation

Existence of eigenvalues and eigenvectors

- Let A be a real symmetric matrix of order $n \geq 2$
- The characteristic polynomial equation

$$p(\lambda) = |A - \lambda I| = 0 \tag{14}$$

of degree n must have at least one solution, say, α

- Then, there is atleast one real eigenvector that lies in the null space of $(A - \lambda I)$ or the kernel of $(A - \lambda I)$

A factorization of p(λ)

- In general, the monic polynomial can be expressed as

$$p(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i} \tag{15}$$

  where n_i is the algebraic multiplicity of λ_i and $(n_1 + n_2 + \ldots + n_k) = n$
- The number of distinct eigenvectors m_i corresponding to a given eigenvalue λ_i is called the geometric multiplicity
- In general, $1 \leq m_i \leq n_i$, when A is symmetric, $m_i = n_i$ for $1 \leq i \leq k$

Claim 1: Eigenvalues of a real symmetric matrix are real

- Let (λ, v) be an eigenpair of A. That is

$$Av = v\lambda \tag{16}$$

- Taking complex conjugates of both sides:

$$A\bar{v} = \bar{v}\bar{\lambda} \tag{17}$$

- Multiplying both sides of (16) by $\bar{v}^T$ on the left and that of (17) by v^T on the left and subtracting

$$0 = \bar{v}^T Av - v^T A\bar{v} = \lambda\bar{v}^T v - \bar{\lambda} v^T \bar{v} = v^T \bar{v}(\lambda - \bar{\lambda}) \tag{18}$$

- Since $v^T\bar{v} > 0, \implies \lambda = \bar{\lambda}$ and hence the claim

Claim 2: Eigenvectors corresponding to different eigenvalues of a real symmetric matrix are orthogonal

- Let (λ, v) and (μ, u) be two eigenpairs of a symmetric matrix A and let $\lambda \neq \mu$

- $$\textit{Then} \quad Av = \lambda v \quad \textit{and} \quad Au = \mu u \tag{19}$$

- Multiplying both sides of the first equation on the left by u^T and that of the second by v^T and subtracting:

$$0 = u^T Av - v^T Au = \lambda u^T v - \mu v^T u = (\lambda - \mu) u^T v \tag{20}$$

- Since $\lambda \neq \mu$, it is immediate that $u^T v = 0$ and the claim follows

A notation

- Let $D \subset R^n$ denote an A-invariant subspace of A. That is, $Av \in D$ when $v \in D$
- Let $D^\perp$ denote the subspace of R^n that is orthogonal to D. That is $u^T v{=}0$ whenever $u \in D$ and $v \in D^\perp$

- For any $u, v \in R^n$

$$v^T Au = (Av)^T u \tag{21}$$

- If $u \in D$, then $Au \in D$. If $v \in D^\perp$, then $v^T Au = 0$
- From $(Av)^T u = 0$, it follows that $Av \in D^\perp$, and the claim is true.

Claim 4: Every (non-null) A-invariant subspace D of A contains a real eigenvector of A

- Let k be the dimension of D. Then there exists a $n \times k$ matrix B whose columns constitute an orthogonal basis for D.
- Since D is A-invariant, it is immediate that

$$AB = BE \tag{22}$$

  for some $E \in R^{k \times k}$
- Then,

$$B^T AB = B^T BE = E \tag{23}$$

  where E is a real symmetric matrix

Proof of claim 4 (Continues)

- Since E is real and symmetric, there exists atleast one eigenpair (λ, x) for E: $Ex = \lambda x$ where $x \in R^k$
- Then $(AB)x = A(Bx) = (BE)x = B(Ex) = \lambda(Bx)$
- Since $x \neq 0$ and the columns of B are orthogonal and hence linearly independent, it follows that $Bx \neq 0$
- Hence, Bx is an eigenvector A contained in D

Claim 5: The set of all n eigenvectors of a real symmetric matrix $A \in R^{n \times n}$ form an orthogonal basis for R^n

- Recall that every real symmetric matrix A is endowed with at least one eigen pair
- Hence, for some $m \geq 1$, let $\{v_1, v_2, \ldots, v_m\}$ be the (orthonormal) eigenvector basis for a subspace D of R^n
- Clearly, D and $D^\perp$ are A-invariant.Hence, there is a vector $v_{m+1} \in D^\perp$ such that $\{v_1, v_2, \ldots, v_{m+1}\}$ are the eigenvectors of A.
- Starting with m=1 and using this inductive argument, we obtain an orthonormal basis for R^n which are eigenvectors of A

Claim 6: Every real symmetric matrix A is diagonalizable

- Given A, let v=$[v_1, v_2, \ldots, v_n] \in R^{n\times n}$ be the matrix of eigenvectors of A, that is $Av_i = v_i\lambda_i$ and $\Lambda = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n) \in R^{n\times n}$
- Then $AV = V\Lambda$ and $V^TV = VV^T = I$
- Hence, $V^TAV = \Lambda$

References

- Appendix follows the developments in Chapter 8 of C. Godsil and G. Royle (2001) Algebraic Graph Theory, Springer Verlag
- G.Golub and C. Van Loan (1989) Matrix Computations, Johns Hopkins University Press contains a wealth of information on computing the eigen pairs of real matrices

MODULE 1.2
Singular Value Decomposition (SVD)

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

What is SVD?

- This module 1.1 contains results relating to the spectral decomposition of square, real symmetric matrices
- This module 1.2 contains analogous results for rectangular matrices, $H \in R^{m \times n}$ called SVD of H
- SVD rests on the spectral decomposition of symmetric matrices $H^T H$ and HH^T are called the Gramian of H

Gramians of H

- Given $H \in R^{m\times n}$, define two related square, symmetric matrices: $H^T H \in R^{n\times n}$ and $HH^T \in R^{m\times m}$ called the Gramians of H
- Assume that H is of full rank, that is,

$$RANK(H) = min(n, m) \tag{1}$$

- From

$$RANK(H^T H) = RANK(H) = RANK(HH^T) \tag{2}$$

it follows that

$$RANK(H^T H) = RANK(HH^T) = min(n, m) \tag{3}$$

- Hence, when $m > n$, $H^T H \in R^{n\times n}$ is non singular and in fact, is SPD. But HH^T is singular and non-negative definite

Spectral decomposition of $H^T H \in R^{n \times n}$ when $m > n$

- Since the smaller Gramian $H^T H$ is an SPD matrix, there exists eigenpairs (λ_i, v_i) $1 \leq i \leq n$ such that

$$(H^T H)V = V\Lambda \tag{4}$$

where $V = [v_1, v_2, \ldots, v_n] \in R^{n \times n}$ and $\Lambda = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n) \in R^{n \times n}$ and $V^T V = VV^T = I_n$

- Also, assume that

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0 \tag{5}$$

- Hence,

$$V^T (H^T H)V = \Lambda \quad \text{and} \quad H^T H = V\Lambda V^T \tag{6}$$

Eigenpair of $HH^T \in R^{m \times m}, m > n$

- Define

$$u_i = \frac{1}{\sqrt{\lambda_i}} Hv_i \in R^m, 1 \le i \le n \tag{7}$$

- Then

$$(H^T H)u_i = \frac{1}{\sqrt{\lambda_i}} H(H^T H)v_i = \frac{\lambda_i}{\sqrt{\lambda_i}} Hv_i = \lambda_i u_i \tag{8}$$

- That is, if (λ_i, v_i) is an eigenpair of $(H^T H)$, then (λ_i, u_i) is an eigenpair of HH^T with u_i given by (7)

Spectral decomposition of $HH^T \in R^{m\times m}, m > n$

- Let $U = [u_1, u_2, \ldots, u_n] \in R^{m\times n}$. Then (7) is equivalent to

$$(HH^T)U = U\Lambda \tag{9}$$

- The n non-zero eigenvalues of (HH^T) are the same as the n eigenvalues of $H^T H$. The rest of the (m-n) eigenvalues of HH^T are zero
- The eigenvectors u_i corresponding to the n non-zero eigenvalues of (HH^T) are related to those of $(H^T H)$ through the linear transformation in (7)

- Relation (7) becomes

$$Hv_i = u_i\sqrt{\lambda_i}, \ \ 1 \le i \le n \tag{10}$$

- Define

$$U = [u_1, u_2, u_3, \ldots, u_n] \quad \in R^{m\times n}$$

$$\Lambda^{\frac{1}{2}} = Diag(\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \ldots, \lambda_n^{\frac{1}{2}}) \quad \in R^{n\times n}$$

- The n relations in (10) can be written succinctly as

$$HV = U\lambda^{\frac{1}{2}} \quad or \quad H = U\lambda^{\frac{1}{2}}V^T \tag{11}$$

called the SVD of H

Has a sum of rank-1 matrices

- Equation(11) on expanding:

$$H = \left[u_1, u_2, u_3, \dots, u_n\right] \begin{bmatrix} \lambda_1^{\frac{1}{2}} & 0 & \dots & 0 \\ 0 & \lambda_2^{\frac{1}{2}} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} \qquad (12)$$

$= \sum_{i=1}^{n} \lambda_i^{\frac{1}{2}} u_i v_i^T$

- λ_i 's are the eigenvalues of $(H^T H)$ and are known as the singular values of H
- Hence the name SVD

A dual pair for SVD

- Multiplying both sides of (7) on the left by H^T and using (7):

$$H^T u_i = \frac{1}{\sqrt{\lambda_i}}(H^T H)v_i = \frac{1}{\sqrt{\lambda_i}}\lambda_i v_i = \sqrt{\lambda_i} v_i$$

- That is,

$$\begin{aligned} v_i &= \frac{1}{\sqrt{\lambda_i}} H^T u_i \;\; and \\ u_i &= \frac{1}{\sqrt{\lambda_i}} H v_i \end{aligned} \tag{13}$$

  are the two defining relations for SVD of H

A Generalization

- Let $\lambda \neq 0, \eta \neq 0$ be such that (λ, η) is an eigenpair of $H^T H$. That is
$$(H^T H)\eta = \lambda \eta \tag{14}$$
- From
$$\lambda(H\eta) = H(H^T H)\eta = (HH^T)(H\eta) \tag{15}$$
it follows that $(\lambda, H\eta)$ is an eigenpair of HH^T
- If $H\eta = 0$, then $(H^T H)\eta = \lambda\eta = 0$ which implies either $\lambda = 0$ or $\eta = 0$ or both zero, which is a contradiction.
- Hence $(\lambda, H\eta)$ is an eigenpair of (HH^T) if (λ, η) is that of $H^T H$

Algebraic and geometric multiplicities of eigenvalues of $H^T H$

- Let λ be an eigenvalue of $(H^T H)$ of algebraic multiplicity, say, m.
- Then, recall that there exists a (non- unique) set of m orthonormal eigenvectors $\{\eta_1, \eta_2, \eta_3, \ldots, \eta_m\}$ such that

$$(H^T H)\eta_i = \lambda \eta_i \quad for \quad 1 \leq i \leq m \tag{16}$$

Algebraic and geometric multiplicities of eigenvalues of HH^T

- Let η_1 and η_2 be two orthogonal eigenvectors of $H^T H$ for the eigenvalue λ of algebraic multiplicity m $= 2$
- Then, $H\eta_1$ and $H\eta_2$ as eigenvectors of (HH^T) are orthogonal
- For

$$(H\eta_1)^T(H\eta_2) = \eta_1^T(H^T H)\eta_2 = \lambda\eta_1^T\eta_2 = 0 \tag{17}$$

- In view of (15) and (17), the following claim holds:
- Claim: Let H be an $m \times n$ matrix of full rank.
  Then
  (1) The Gramians $H^T H$ and HH^T share the same set of non-zero eigenvalues, and
  (2) λ is an eigenvalue of multiplicity m of $(H^T H)$ with an orthogonal set of eigenvectors $\{\eta_1, \eta_2, \eta_3, \ldots, \eta_m\}$, then λ is also an eigenvalue of multiplicity m of (HH^T) with an orthogonal set of eigenvectors $\{H\eta_1, H\eta_2, \ldots, H\eta_m\}$

Spectral decomposition of $HH^T \in R^{m \times m}, n > m$

- For completeness, we consider the case when $n > m$
- Since (HH^T) is SPD, there exist (λ_i, u_i), $1 \leq i \leq n$ that are eigenpairs of HH^T
- That is,

$$\begin{aligned} (HH^T u_i) &= \lambda_i u_i, \quad u_i \in R^n \\ or \quad (HH^T)U &= U\lambda \end{aligned} \tag{18}$$

where $U = [u_1, u_2, u_3, \ldots, u_n], U^T U = UU^T = I_m$
$\Lambda = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$
where

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \tag{19}$$

- Define

$$v_i = \frac{1}{\sqrt{\lambda_i}} H^T u_i \in R^n \tag{20}$$

- Then

$$(H^T H)v_i = \frac{1}{\sqrt{\lambda_i}} H^T (HH^T)u_i = \frac{\lambda_i}{\sqrt{\lambda_i}} H^T u_i = \lambda_i v_i \tag{21}$$

- That is, (λ_i, v_i) is an eigenpair of $H^T H$

Eigen decomposition of $H^T H$

- Define

$$V = [v_1, v_2, v_3, \ldots, v_n] \in R^{n \times n}$$

- Then (21) becomes

$$(H^T H)V = V\Lambda, \quad vv^T = I_n \tag{22}$$

- Also the m non-zero eigenvalues of HH^T are those of HH^T and the rest of (n-m) eigenvalues of $H^T H$ are zero.

Dual of (20)

- Multiplying both sides of (20) on the left by H and using (18):

$$Hv_i = \frac{1}{\sqrt{\lambda_i}}(HH^T)u_i = \sqrt{\lambda_i}u_i$$

$$or \quad u_i = \frac{1}{\sqrt{\lambda_i}}Hv_i \tag{23}$$

which is the dual of (20)

A note on our notation

- In this and in all Modules to follow, we use the following convention: $H \in R^{m \times n}$
- Case 1: $m > n$ and $H^T H$ is SPD

$$\begin{aligned} (H^T H)V = V\Lambda, \quad V^T V = VV^T = I_n \\ (HH^T)U = U\Lambda, \quad U^T U = I_n, \quad U \in R^{m \times n} \end{aligned} \tag{24}$$

- Case 2: $n > m$ and HH^T is SPD

$$\begin{aligned} (HH^T)U = U\Lambda, \quad U^T U = UU^T = I_m \\ (H^T H)V = V\Lambda, \quad V^T V = I_m, \quad V \in R^{n \times m} \end{aligned} \tag{25}$$

A dual characterization of SVD

- Case 1: $m > n$

$$
\begin{aligned}
u_i &= \frac{1}{\sqrt{\lambda_i}} H v_i \\
v_i &= \frac{1}{\sqrt{\lambda_i}} H^T v_i
\end{aligned}
\tag{26}
$$

- Case 2: $n > m$

$$
\begin{aligned}
v_i &= \frac{1}{\sqrt{\lambda_i}} H^T u_i \\
u_i &= \frac{1}{\sqrt{\lambda_i}} H v_i
\end{aligned}
\tag{27}
$$

MODULE 1.3
Orthogonal Projections in R^m

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

Inner product and norm in R^m

- Let $x, y \in R^m$. The inner product is defined as

$$< x, y >= x^T y = \sum_{i=1}^{n} x_i y_i = y^T x =< y, x > \quad (1)$$

- Norm of a vector x is given by

$$||x|| =< x, x >^{\frac{1}{2}} = (\sum_{i=1}^{m} x_i^2)^{\frac{1}{2}} \quad (2)$$

- Cauchy- Schwartz inequality: From

$$< x, y >= ||x|| \, ||y|| \, cos(\theta) \quad (3)$$

it follows that

$$|cos\theta| = \frac{| < x, y > |}{||x|| \, ||y||} \leq 1 \quad (4)$$

- Let $\hat{x} = \frac{x}{||x||}$ be the unit vector,

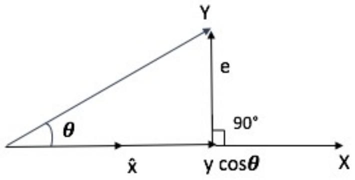
$$||\hat{x}|| = 1 \tag{5}$$

-

$$< y, \hat{x} >= y^T \hat{x} = ||y|| cos(\theta) \tag{6}$$

which is the component of y in the direction $\hat{x}$.

- The vector $(y cos\theta)\hat{x}$ is called the projection of y along $\hat{x}$



Orthogonality of this projection

- Let $e = y - (ycos\theta)\hat{x}$ be the error in the projection
- Then,

$$< e, \hat{x} >= e^T\hat{x} = y^T\hat{x} - (ycos\theta)\hat{x}^T\hat{x} = 0 \tag{7}$$

- Hence the name orthogonal projection

Analytical expression for orthogonal projection

- Let $h \in R^m$ and $x \in R^m$ be any other vector
- Any vector along h can be expressed as a multiple $h\alpha$ for some real α
- Problem: Given x and h, find $\alpha \in R$ that minimizes the distance between x and $h\alpha$
- That is, find α that minimizes

$$
\begin{aligned}
Q(\alpha) = ||x - h\alpha||^2 &= (x - h\alpha)^T(x - h\alpha) \\
&= x^T x - 2x^T h\alpha + \alpha^2 h^T h \qquad (8)
\end{aligned}
$$

Optimal α

- Minimizer α^* is obtained by solving

$$0 = \frac{dQ}{d\alpha} = -2h^T x + 2\alpha h^T h \tag{9}$$

- That is,

$$\alpha^* = (h^T h)^{-1} h^T x = h^+ x \tag{10}$$

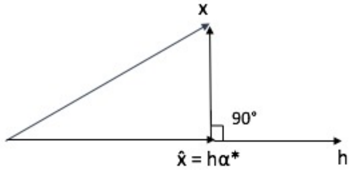
where

$$h^+ = (h^T h)^{-1} h^T \tag{11}$$

is called the generalized inverse of h

Expression for the projection

- The orthogonal projection of x along h is given by



$$\hat{x} = h\alpha^* = hh^+x = h(h^Th)^{-1}h^Tx = P_hx \qquad (12)$$

where

$$P_h = hh^+ = h(h^Th)^{-1}h^T \qquad (13)$$

is called the orthogonal projection matrix

Orthogonality of projection

- Let
$$e = x - \hat{x} = (I - P_h)x \tag{14}$$
be the error in the projection
- Clearly:
$$h^T e = (h^T - h^T P_h)x = 0 \tag{15}$$
since $h^T P_h = (h^T h)(h^T h)^{-1} h^T = h^T$
- Hence, P_h is called the orthogonal projection operator

Properties of P_h

- Symmetry: $P_h^T = P_h$
- Idempotent: $P_h^2 = P_h$
- P_h is a rank one matrix
- $det(P_h) = 0$, that is, P_h is singular
- 1 is the only non-zero eigenvalue of P_h
- P_h is not an orthogonal matrix: $P_h^T \neq P_h^{-1}$ since P_h^{-1} is not defined

A Generalization

- Let $H \in R^{m \times n}$ with m>n and Rank(H) = n
- Then, $(H^T H) \in R^{n \times n}$ is SPD
- Let $x \in R^m$
- Problem: Find an $\alpha \in R^n$ such that $\hat{x} = H\alpha \in R$ and

$$\begin{aligned} Q(\alpha) &= (x - H\alpha)^T (x - H\alpha) = ||x - H\alpha||^2 \\ &= x^T x - 2x^T H\alpha + \alpha^T (H^T H)\alpha \end{aligned} \tag{16}$$

is a minimum

Optimal α

- From

$$\nabla_\alpha Q(\alpha) = -2H^T x + 2(H^T H)\alpha = 0 \tag{17}$$

it follows that

$$\alpha^* = (H^T H)^{-1} H^T x = H^+ x \tag{18}$$

minimizes $Q(\alpha)$ since

$$\nabla_\alpha^2 Q(\alpha) = (H^T H) \quad is \quad SPD \tag{19}$$

- $H^+ = (H^T H)^{-1} H^T \in R^{n \times m}$ is called the generalized inverse of H

- Then

$$\hat{x} = H\alpha^* = H(H^TH)^{-1}H^Tx = HH^+x = P_Hx \tag{20}$$

where

$$P_H = H(H^TH)^{-1}H^T \in R^{m\times m} \tag{21}$$

is the projection operator in R^m onto the n-dimensional subspace spanned by the columns of H

- $$e = x - \hat{x} = (I - P_H)x \tag{22}$$

is the error in this projection

- Verify that

$$e^TH = 0 \tag{23}$$

and hence the name orthogonal projection

Properties of P_H

- Symmetry: $P_H^T = P_H$
- Idempotent: $P_H^2 = P_H$
- $RANK(P_H) = n$ since that of H is n
- P_H is singular
- There are exactly n non-zero eigenvalues of P_H
- P_H is not an orthogonal matrix

Reference

- Chapters 5 and 6 in J. Lewis, S. Lakshmivarahan and S.K. Dhall (2006) Dynamic Data Assimilation, Cambridge University Press.

Exercises

1) Let $x \in R^m$ and $h = (1, 1, \ldots, 1)^T \in R^m$ and $h = (1, 1, 1, \ldots, 1)^T \in R^m$ be a vector all of whose components are 1. Compute an expression for α that minimizes the distance between x and $h\alpha$.

2) Let H $= \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$

- Compute $H^T H$, HH^T, H^+, P_H, HH^+, H^+H, HH^+H, H^+HH^+
- Compute the eigenvalues of P_H

3) Verify the following:
   a) $HH^+H = H$
   b) $H^+HH^+ = H^+$
   c) $(H^+H)^T = H^+H$
   d) $(HH^+)^T = HH^+$
   Note: Any H^+ satisfying the properties (a)-(d) is called the Moore-Penrose inverse.

4) Given P_H, define $P_H^{\perp} = I - P_H$
   - Verify that $P_H^{\perp}$ is symmetric and idempotent
   - For the H in problem 2, Compute $P_H^{\perp}$ and its rank

MODULE 1.5
Second-order properties of random variables and vectors

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

L_2 - space of square integrable random variables

- Let (Ω, Γ, P) be a probability space
- $L_2 = L_2(\Omega, \Gamma, P)$ denote the family of square integrable random variables
- Say that $x(\omega) \in L_2$ if

$$\int_{\Omega} |x(\omega)|^2 dP(\omega) < \infty \tag{1}$$

- L_2 is a real vector space-closed under addition and multiplication by a real constant

Second-order properties of random variables

- Let $x \in L_2$, with mean $\mu_x = E[x] < \infty$
- Variance of x: $var(x) = \sigma_x^2 = E[(x-\mu)^2] < \infty$
- Covariance between $x, y \in L_2$:

$$cov(x,y) = E[(x-\mu_x)(y-\mu_y)]$$

- x, y are uncorrelated if cov(x, y)$=0$
- Correlation between $x, y \in L_2$:

$$corr(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

- $|corr(x,y)| \leq 1$

A Geometric view of random variables in L_2

- Let $x, y \in L_2$
- Inner product of x and y: $< x, y >= E(xy)$
- Norm of x: $||x|| =< x, x >^{\frac{1}{2}}= [E(x^2)]^{\frac{1}{2}}$
- Distance between x and y:

$$dist(x, y) = ||x - y|| = [E(x - y)^2]^{\frac{1}{2}}$$

- x and y are orthogonal if $< x, y >= E(x, y) = 0$
- For mean zero random variables: orthogonality implies uncorrelated

Orthogonal projection in L_2

- Let x be a random variable defined on (Ω, Γ, P)
- Let y be a random variable on a subspace (Ω, Y, P) where Y is a sub σ-field of Γ
- Orthogonal projection theorem: For $x \in L_2(\Omega, \Gamma, P)$ there exists an unique $\hat{x} \in L_2(\Omega, Y, P)$ such that
  $(a) ||x - \hat{x}|| = min\{||x - y|| : y \in L_2(\Omega, Y, P)\}$
  $(b) < x - \hat{x}, y > = 0$ for all $y \in L_2(\Omega, Y, P)$

Random vectors in L_2

- Let $x \in R^m$ be a random vector with $x = (x_1, x_2, x_3, \ldots, x_m)$
- Say $x \in L_2$ if each component $x_i \in L_2$
- Mean of $x = \mu_x = E(x) = (\mu_1, \mu_2, \mu_3, \ldots, \mu_m)^T$ where $\mu_i = E(x_i)$
- $cov(x_i, x_j) = \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$
- $var(x_i) = \sigma_i^2 = E[(x_i - \mu_i)^2]$
- $cov(x) = E[(x - \mu)(x - \mu)^T] = [\sigma_{ij}] = \Sigma \in R^{m \times m}$
- $var(x) = \sum_{i=1}^{n} var(x_i)$
  $= E[(x - \mu)^T (x - \mu)] = \sum_{i=1}^{n} \sigma_i^2 = tr(\Sigma)$
  where $tr(A)$ is called the trace of A.

A geometric view of random vectors in L_2

- Let $x, y \in R^m$ be two random vectors in L_2
- Inner product: $< x, y >= E[x^T y] = \sum_{i=1}^{m} E(x_i y_i)$
- Norm: $||x||^2 =< x, x >= E[x^T x] = \sum_{i=1}^{m} E(x_i)^2$
- Distance: $||x - y||^2 =< x - y, x - y >$
  $= E[(x - y)^T (x - y)] = \sum_{i=1}^{m} E(x_i - y_i)^2$
- Orthogonal: x and y are orthogonal if $< x, y >= 0$
- For mean zero random vectors orthogonality implies uncorrelated

Orthogonal projection

- The statement of orthogonal projection theorem carries over verbatim if we replace random variables by random vectors
- This projection theorem is the basis for generating optimal prediction, optimal estimation in Time Series Analysis, Spatial and Spatio-temporal statistics.
- It also plays a key role in the Principal Component Analysis (PCA) and in the development of Empirical orthogonal functions (EOF)

Centering

- Let $x \in R^m$ be a random vector with mean $\mu \in R^m$ and $cov(x) = \Sigma \in R^{m \times m}$
- Then, $y = x - \mu$ is called the centered version of x
- Clearly: E(y) = 0 and $cov(y) = \Sigma$

Normalization of $x \in R^n$

- Let $x \in R^m$ be a random vector with mean $\mu \in R^m$ and $cov(x) = \Sigma$ with $y = x - \mu$
- $z_i = \frac{x_i - \mu_i}{\sigma_i} = \frac{y_i}{\sigma_i}$ is the normalized version of y_i
- Mean$(z_i) = 0$, Var$(z_i) = 1$
- $z = (z_1, z_2, \ldots, z_m)^T$ is the centered and normalized version of x

Normalization (continued)

- Let $D = Diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ - the diagonal matrix with the diagonal of Σ
- Define square root, $D^{1/2} : D = D^{1/2}D^{1/2}$ where $D^{1/2} = Diag(\sigma_1, \sigma_2, \ldots, \sigma_m)$
- Define $z = D^{-1/2}Y = D^{-1/2}(x - \mu)$
- $cov(z) = E(zz^T) = D^{-1/2}E[(x-\mu)(x-\mu)^T]D^{-1/2}$ $= D^{-1/2}\Sigma D^{-1/2} = R = corr(z)$
- Correlation matrix: $R = [R_{ij}]$ and $R_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$
- $|R_{ij}| \leq 1$

Linear Transformation of $x \in R^m$

- Let $A \in R^{m \times m}$ and $b \in R^m$
- Define $\xi = Ax + b$
- Mean: $E(\xi) = A\mu + b$ where $\mu = E(x)$
- $$\begin{aligned} cov(\xi) &= E[(\xi - E(\xi))(\xi - E(\xi))^T] \\ &= E[(A(x-\mu))(A(x-\mu))^T] \\ &= AE[(x-\mu)(x-\mu)^T]A^T = A\Sigma A^T \end{aligned}$$
- Thus, if $x \sim N(m, \Sigma)$, $\xi \sim N(Am + b, A\Sigma A^T)$

A special linear transformation

- Let $x \in R^m$ with mean μ and $cov(x) = \Sigma$, SPD
- Define square root of $\Sigma : \Sigma = \Sigma^{1/2}\Sigma^{1/2}$
- let $\xi = \Sigma^{-1/2}(x - \mu)$
- Then $E(\xi) = 0$
- $$\begin{aligned} cov(\xi) &= E[(\Sigma^{-1/2}(x-\mu))(\Sigma^{-1/2}(x-\mu))^T] \\ &= \Sigma^{-1/2}E[(x-\mu)(x-\mu)^T]\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I \end{aligned}$$
- That is: $var(\xi_i) = 1$ and $cov(\xi_i, \xi_j) = 0$ for i $\neq$ j
- This is known as Whitening transformation

Linear functional of x

- Let $a \in R^m$ and $x \in R^m$ with mean μ and $cov(x)$
- Define $\eta = a^T x$, a real random variable
- $E(\eta) = a^T \mu$
- $$\begin{aligned} var(\eta) &= E[(a^T(x-\mu))^2] = E[(a^T(x-\mu))(a^T(x-\mu))] \\ &= E[a^T(x-\mu)(x-\mu)^T a^T] \\ &= a^T \Sigma a \end{aligned}$$
- Clearly, η is a non-degenerate random variable (that is, var(η) $>$0) for all $a \in R^n$, if and only if Σ is SPD

- M.Grigoriu (2002) Stochastic Calculus, Brikhauser, Basel contains a good introduction to basic Probability theory and L_2 spaces

Exercise

1. Prove that $|R_{ij}| = |\frac{\sigma_{ij}}{\sigma_i \sigma_j}|$