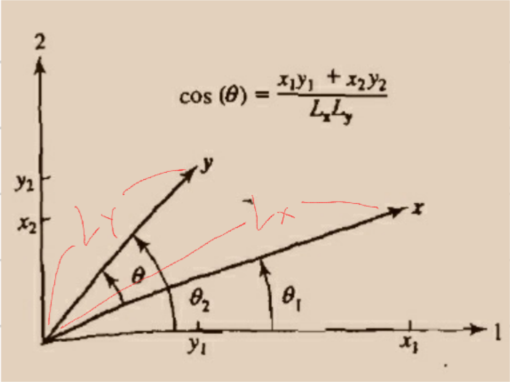


$$\cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_x L_y}$$

$$\cos(\theta_1) = \frac{x_1}{L_x}, \quad \cos(\theta_2) = \frac{y_1}{L_y}$$

$$\sin(\theta_1) = \frac{x_2}{L_x}, \quad \sin(\theta_2) = \frac{y_2}{L_y}$$

Length of a vector

$$L_x = \sqrt{X'X} \Rightarrow \cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_x L_y} = \frac{X'Y}{\sqrt{X'X}\sqrt{Y'Y}} = \frac{X'Y}{L_x L_y}$$

※ at $\cos(90x)$, $x = 1, 2, 3 \dots$, $X'Y = 0$ because of orthogonality

Linear Dependency :

There is one or more nonzero constants satisfying

$$c_1 X + c_2 Y = 0,$$

and these constants are called "solutions".

Projection of X on Y :

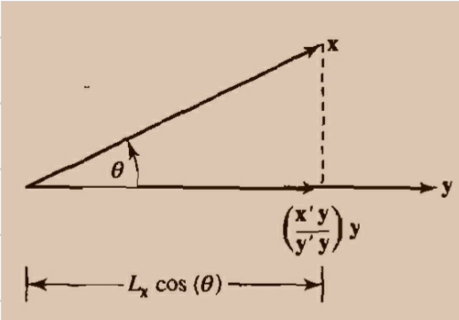
"is projection the orthogonal point of X on Y?"
- Y에서 X 까지 제일 가까운 수직의 점...?"

$$\frac{X'Y}{Y'Y} Y = \frac{X'Y}{L_y} \frac{1}{L_y} Y, \quad L(L_y^{-1} Y) = 1$$

※ $L_y = \sqrt{Y'Y}$

Length of the projection :

$$\frac{|X'Y|}{L_y} = L_x \left| \frac{X'Y}{L_x L_y} \right| = L_x |\cos(\theta)|$$



Orthogonal Matrix :

$$QQ' = Q'Q = I, \quad \text{or} \quad Q' = Q^{-1}$$

Eigenvalues :

$$Ax = \lambda x$$

2.3 Positive Definite Matrices

Spectral Decomposition : an expansion for symmetric matrices

$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots \lambda_k e_k e_k'$, e_i = normalized eigenvectors

※ e_i's are the normalized solutions of the equations $Ae_i = \lambda_i e_i$

- Spectral decomposition can demonstrate important statistical results.
  One is "matrix explanation of distance."

Since A is a symmetric matrix, $x'Ax$ only has squared terms x_i^2 and $x_i x_j$, and it is called quadratic form, and if $0 \le x'Ax$, the symmetric matrix A and the quadratic form are called nonnegative definite. If $0 = x'Ax$ holds only for $x' = 0$, then A or the quadratic form are called positive definite.

See EX 2.11 in pdf 84 for further explanation

- find λ's for $|A - \lambda I| = 0$
- decompose A to,

$$A_{m\times m} = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots \lambda_m e_m e_m'$$

e_i's are normalized and orthogonal eigenvectors associated with each λ_i's

$$x'Ax = \lambda_1 x' e_1 e_1' x + \lambda_2 x' e_2 e_2' x + \cdots \lambda_m x' e_m e_m' x$$

note $x' e_i = e_i' x$

setting $y_i = x' e_i = e_i' x$, show y_i's are nonzero and therefore A is positive definite. From, $x'Ax = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots$, we have

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} e_1' \\ \vdots \\ e_m' \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \Rightarrow Y = E X, \quad E = \text{orthogonal matrix}$$

$X = E'Y$, because X is nonzero vector, the equation implies $Y \neq 0$

using the spectral decomposition, we can easily show that a $k\times k$ symmetric matrix A is a positive definite matrix iff every eigenvalue of $A > 0$

※ Symmetric matrices are positive definite matrices iif $\lambda_i > 0$

" nonnegative " $\lambda_i \ge 0$

should, be interpreted in terms of standard deviation units. In this way, we can account for the inherent uncertainty (variability) in the observations. Points with the same associated "uncertainty" are regarded as being at the same distance from the origin. = points lying at the tip of an ellipse

$\sum \lambda_i (x'e_i)^2$

$(\text{distance})^2 = x'Ax = \sum \lambda_i x' e_i e_i' x$ 이라는 건가?

Pdf 85: In sum, distance is determined from a positive definite quadratic form $x'Ax$. Conversely, a positive definite quadratic form can be interpreted as a squared distance.

As mentioned earlier, the statistical distance from the origin can be shown,

$$d^2(O,P) = a_{11}x_1^2 + a_{22}x_2^2 + \cdots a_{pp}x_p^2 + 2(a_{12}x_1x_2 + \cdots + a_{p-1}x_{p-1}x_p),$$

and this equation can derive,

$$d^2(O,P) = [x_1 \ x_2 \cdots x_p] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{p1} & \cdots & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = x'Ax,$$

from there, notice that for $x \neq 0$, symmetric matrices are positive definite.

Comment. Let the square of the distance from the point $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ to the origin be given by $\mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a $p \times p$ symmetric positive definite matrix. Then the square of the distance from $\mathbf{x}$ to an arbitrary fixed point $\boldsymbol{\mu}' = [\mu_1, \mu_2, \ldots, \mu_p]$ is given by the general expression $(\mathbf{x} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{x} - \boldsymbol{\mu})$.

Example from the text book

Expressing distance as the square root of a positive definite quadratic form allows us to give a geometrical interpretation based on the eigenvalues and eigenvectors of the matrix $\mathbf{A}$. For example, suppose $p = 2$. Then the points $\mathbf{x}' = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 = c^2$$

By the spectral decomposition, as in Example 2.11,

$$\mathbf{A} = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2' \quad \text{so} \quad \mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2(\mathbf{x}'\mathbf{e}_2)^2$$

Now, $c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is an ellipse in $y_1 = \mathbf{x}'\mathbf{e}_1$ and $y_2 = \mathbf{x}'\mathbf{e}_2$ because $\lambda_1, \lambda_2 > 0$ when $\mathbf{A}$ is positive definite. (See Exercise 2.17.) We easily verify that $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$ satisfies $\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(c\lambda_1^{-1/2}\mathbf{e}_1'\mathbf{e}_1)^2 = c^2$. Similarly, $\mathbf{x} = c\lambda_2^{-1/2}\mathbf{e}_2$ gives the appropriate distance in the $\mathbf{e}_2$ direction. Thus, the points at distance c lie on an ellipse whose axes are given by the eigenvectors of $\mathbf{A}$ with lengths proportional to the reciprocals of the square roots of the eigenvalues. The constant of proportionality is c. The situation is illustrated in Figure 2.6.

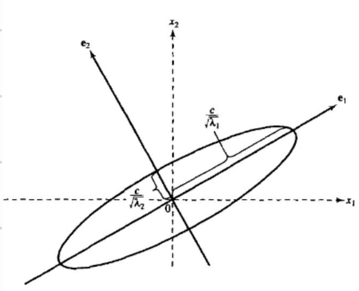


Figure 2.6 Points a constant distance c from the origin ($p = 2, 1 \leq \lambda_1 < \lambda_2$).

If $p > 2$, the points $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ a constant distance $\boxed{c = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}}$ from the origin lie on hyperellipsoids $c^2 = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \cdots + \lambda_p(\mathbf{x}'\mathbf{e}_p)^2$, whose axes are given by the eigenvectors of $\mathbf{A}$. The half-length in the direction $\mathbf{e}_i$ is equal to $c/\sqrt{\lambda_i}$, $i = 1, 2, \ldots, p$, where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{A}$.

2.4 Square-Root Matrix

Let A be $k \times k$ positive definite matrix, then A's spectral decomposition is,

$A = \sum_{i=1}^{k} \lambda_i e_i e_i'$, and assume we have a symmetric matrix P, which $P = [e_1\ e_2 \cdots e_k]$, then the following equation is verified.

$$A = \sum_{i=1}^{k} \lambda_i e_i e_i' = \underbrace{P \Lambda P'}_{\text{singular value decomp.}}, \text{ where } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_k \end{bmatrix}$$

$$PP' = P'P = I$$

$$= \operatorname{diag}[\lambda_1, \lambda_2, \ldots, \lambda_k]$$

From the above assumptions and equations, we can derive that,

$$A^{-1} = P \Lambda^{-1} P' = \sum_{i=1}^{k} \frac{1}{\lambda_i} e_i e_i', \text{ and let } \Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sqrt{\lambda_k} \end{bmatrix}$$

then $\sum_{i=1}^{k} \sqrt{\lambda_i}\, e_i e_i' = P \Lambda^{\frac{1}{2}} P'$ is called the "square-root of A." $= A^{\frac{1}{2}}$

Properties of Square-root matrix

1. $(A^{\frac{1}{2}})' = A^{\frac{1}{2}}$
2. $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$.
3. $(A^{\frac{1}{2}})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} e_i e_i' = P \Lambda^{-\frac{1}{2}} P'$, $\Lambda^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \frac{1}{\sqrt{\lambda_k}} \end{bmatrix}$
4. $A^{\frac{1}{2}} A^{-\frac{1}{2}} = A^{-\frac{1}{2}} A^{\frac{1}{2}} = I$

   $A^{-\frac{1}{2}} A^{-\frac{1}{2}} = A^{-1}$

2.5 Random Vectors and Matrices

Random Vectors/Matrices :
- Vectors/matrices whose elements are random variables.

Let X and Y be random matrices and A and B be conformable matrices

$E(X+Y) = E(X) + E(Y)$

$E(AXB) = AE(X)B$

2.6 Mean Vectors and Covariance Matrices

Calculating Covariance with joint probability density functions :

$$\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k)$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k)\,dx_i\,dx_k & \text{if } X_i, X_k \text{ are continuous random variables with the joint density function } f_{ik}(x_i, x_k) \\ \displaystyle\sum_{\text{all } x_i}\sum_{\text{all } x_k} (x_i - \mu_i)(x_k - \mu_k) p_{ik}(x_i, x_k) & \text{if } X_i, X_k \text{ are discrete random variables with joint probability function } p_{ik}(x_i, x_k) \end{cases}$$

symmetric

Deriving PXP correlation coefficient matrix from PXP standard deviation matrix

Let $\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$ and $\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$, and

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

, using Σ and $V^{\frac{1}{2}}$,

Final Result

$$\mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2} = \boldsymbol{\Sigma}$$

$$\boldsymbol{\rho} = (\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1}$$

Partitioning the Covariance Matrix

Variable matrix $\mathbb{X}$ can be partitioned into severals, which can be called as characteristics or total collection. Say, $\mathbb{X}$ has two characteristics,

$$\mathbb{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{bmatrix} \begin{matrix} \}\, q \\ \\ \}\, p-q \end{matrix} = \begin{bmatrix} \mathbb{X}^{(1)} \\ \mathbb{X}^{(2)} \end{bmatrix} \quad \text{and} \quad \mu = E(\mathbb{X}) = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \hline \mu_{q+1} \\ \vdots \\ \mu_p \end{bmatrix} \begin{matrix} \}\, q \\ \\ \}\, p-q \end{matrix} = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}$$

$$(\mathbb{X}^{(1)} - \mu^{(1)})(\mathbb{X}^{(2)} - \mu^{(2)}) = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_q - \mu_q \end{bmatrix} \begin{bmatrix} X_{q+1} - \mu_{q+1}, X_{q+2} - \mu_{q+2}, \dots, X_p - \mu_p \end{bmatrix}$$

$$= \begin{bmatrix} (X_1-\mu_1)(X_{q+1}-\mu_{q+1}) & (X_1-\mu_1)(X_{q+2}-\mu_{q+2}) & \cdots & (X_1-\mu_1)(X_p-\mu_p) \\ (X_2-\mu_2)(X_{q+1}-\mu_{q+1}) & (X_2-\mu_2)(X_{q+2}-\mu_{q+2}) & \cdots & (X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_q-\mu_q)(X_{q+1}-\mu_{q+1}) & (X_q-\mu_q)(X_{q+2}-\mu_{q+2}) & \cdots & (X_q-\mu_q)(X_p-\mu_p) \end{bmatrix}$$

$$E(X^1-\mu^1)(X^2-\mu^2) = \begin{bmatrix} \sigma_{1,q+1} & \sigma_{1,q+2} & \cdots & \sigma_{1p} \\ \sigma_{2,q+1} & \sigma_{2,q+2} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{q,q+1} & \sigma_{q,q+2} & \cdots & \sigma_{qp} \end{bmatrix} = \underline{\Sigma_{12}}$$

not necessarily symmetric

* it gives all the covariances σ_{ij}. - Covariances between different collections?

EX:

$(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$

$$= \begin{bmatrix} \underset{(q\times1)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})} \underset{(1\times q)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'} & \underset{(q\times1)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})} \underset{(1\times(p-q))}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'} \\ \underset{((p-q)\times1)}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})} \underset{(1\times q)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'} & \underset{((p-q)\times1)}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})} \underset{(1\times(p-q))}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'} \end{bmatrix}$$

** $\Sigma_{ij} = \Sigma_{ji}'$

and consequently,

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{matrix} q \\ p-q \end{matrix} \overset{\quad q \qquad p-q}{\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}}_{(p\times p)}$$

$$= \left[\begin{array}{ccc|ccc} \sigma_{11} & \cdots & \sigma_{1q} & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ \hline \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pq} & \sigma_{p,q+1} & \cdots & \sigma_{pp} \end{array}\right] \qquad (2\text{-}40)$$

Σ_{12}

Σ_{21}

The Mean Vector and Covariance Matrix for Linear Combinations of Random Variables

$$
\begin{aligned}
Cov(aX_1, bX_2) &= E(aX_1 - a\mu_1)(bX_2 - b\mu_2) \\
&= ab\,E(X_1-\mu_1)(X_2-\mu_2) \\
&= ab\,Cov(X_1, X_2) = ab\,\sigma_{12}
\end{aligned}
$$

$$
\begin{aligned}
Var(aX_1 + bX_2) &= E[(aX_1+bX_2) - (a\mu_1 + b\mu_2)]^2 \\
&= E[a(X_1-\mu_1) + b(X_2-\mu_2)]^2 \\
&= E[a^2(X_1-\mu_1)^2 + b^2(X_2-\mu_2)^2 + 2ab(X_1-\mu_1)(X_2-\mu_2)] \\
&= a^2 Var(X_1) + b^2 Var(X_2) + 2ab\,Cov(X_1, X_2) \\
&= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\,\sigma_{12} \\
&= Var(c'X) \\
&= c'\Sigma c
\end{aligned}
$$

$$\mu_Z = E(Z) = E(CX) = C\mu_X$$

$$\Sigma_Z = Cov(Z) = cov(CX) = C\Sigma_X C'$$

2.7 Matrix Inequalities and Maximization

시험에 안나올거 같은데...?

ppt에 안보임

Cauchy-Schwarz Inequality :

- Let b and d be any $p \times 1$ vectors, then

$(b'd)^2 \le (b'b)(d'd)$, iif $b = cd$ or $d = cb$ for some constant.

Extended Cauchy-Schwarz Inequality :

- let b and d be any two vectors, and let B be a positive definite matrix.

$(b'd)^2 \le (b'Bb)(d'B^{-1}d)$ iif $b = cB^{-1}d$

Maximization Lemma :

- let B be positive definite and d be a given vector. Then for nonzero vector x,

$$\max_{x \ne 0} \frac{(x'd)^2}{x'Bx} = d'B^{-1}d ,$$ when $x = cB^{-1}d$ for any constant.

Maximization of Quadratic Forms for Points on the Unit Sphere.

- let B be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and associated normalized eigenvectors e_j's. Then,

$$\max_{x \neq 0} \frac{x'Bx}{x'x} = \lambda_1 \quad , \text{ when } x = e_1$$

$$\min_{x \neq 0} \frac{x'Bx}{x'x} = \lambda_p \quad , \text{ when } x = e_p$$

$$\max_{x \perp e_1, \ldots, e_k} \frac{x'Bx}{x'x} = \lambda_{k+1} \quad , \text{ when } x = e_{k+1} \,, \; k = 1, 2, \ldots, p-1$$

Supplement 2A - Vectors and Matrices: Basic Concepts

Vectors

Definition of vectors
Definition of scalar multiplication
Definition of vector addition
Definition of vector space
Definition of Linear Combination
Definition of Linear Span
Definition of Linear Independence / Dependence

Basis :
- set of m linearly independent vectors for vector space of all m-tuples
  (tuple : # of entries in a vector)

Definition of Vector Length from the origin

Angle θ between X and Y

$$\cos(\theta) = \frac{(x_1y_1 + x_2y_2 + \cdots + x_my_m)}{L_xL_y} = \frac{X'Y}{\sqrt{X'X}\sqrt{Y'Y}}$$

$$\therefore \; \theta = \cos^{-1}\left\{\frac{x_1y_1 + x_2y_2 + \cdots + x_my_m}{L_x L_y}\right\} = \cos^{-1}\left\{\frac{X'Y}{\sqrt{X'X}\sqrt{Y'Y}}\right\}$$

Definition of Inner Product / Dot Product

Things to be remembered :
1. Z is perpendicular to every vector if and only if $Z = 0$
2. If Z is perpendicular to each vector $X_1, X_2, \ldots, X_K$, then Z is perpendicular to their linear span.
3. Mutually perpendicular vectors are linearly independent

Projection of X on Y :

$$proj(X, Y) = \frac{X'Y}{L_Y^2} Y$$

Gram - Schmidt Process :

Given linearly independent vectors $X_1, \ldots, X_K$, there exist mutually perpendicular vectors $u_1, u_2, \ldots, u_K$ with the same linear span.

$$u_1 = X_1$$

$$u_2 = X_2 - \frac{X_2' u_1}{u_1' u_1} u_1$$

$$\vdots$$

$$u_K = X_K - \frac{X_K' u_1}{u_1' u_1} u_1 - \cdots - \frac{X_K' u_{K-1}}{u_{K-1}' u_{K-1}} u_{K-1}$$

If u_i's are converted to $z_i = \frac{u_i}{\sqrt{u_i' u_i}}$, which has a unit length, projection of X_K on the linear span of $X_1, X_2, \ldots, X_{K-1}$ is,

$$\sum_{i=1}^{K-1} (X_K' z_i) z_i$$

Matrices

Definition of matrix

Definition of dimensions

Definition of equal matrices

Definition of Matrix addition

Definition of matrix scalar multiplication

Definition of matrix subtraction

Definition of transpose

Definition of square matrix

Definition of symmetric matrix

Definition of identify matrix

Definition of matrix multiplication

Definition of determinant

Row rank :
- maximum number of linear independent rows

equal in value

Column rank :
- the rank of its set of columns

"A is linearly independent" = "A is nonsingular"
"A is linearly dependent" = "A is singular"
※ square matrix is nonsingular if its rank is equal to the number of rows
※ If A is nonsingular, there is a unique inverse matrix. $\Rightarrow |A| \neq 0$

Characteristics of inverse matrices
1. $(A^{-1})' = (A')^{-1}$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $|A| = |A'|$
4. $|A| = 0$ if any row is closed under multiplication or addition of the others
5. $|A||A^{-1}| = 1$
6. $|AB| = |A||B|$
7. $|cA| = c^k|A|$, c is a scalar

Definition of Trace

Characteristics of Trace
1. $tr(cA) = c\,tr(A)$
2. $tr(A \pm B) = tr(A) \pm tr(B)$
3. $tr(AB) = tr(BA)$
4. $tr(B^{-1}AB) = tr(A)$
5. $tr(AA') = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}^2$

Characteristics of orthogonal matrices

1. $A^{-1} = A'$
2. $AA' = A'A = I$
3. columns are mutually perpendicular and have unit lengths.

"Eigen values" = "characteristic roots"

Characteristic equation

$\Rightarrow \quad |A - \lambda I| = 0$

ex) A is 2x2

$$|A - \lambda I| = 0$$

$$\begin{vmatrix} a_{11}-\lambda & a_{21} \\ a_{12} & a_{22}-\lambda \end{vmatrix} = 0$$

$(a_{11}-\lambda)(a_{22}-\lambda) - a_{21}a_{12} = 0$, and arrange it in terms of λ.

usually, assuming A is $k \times k$, there are k number of λ's.

Eigenvectors: (characteristic vector)

- one or more nonzero vectors that satisfy the following,

$$Ax = \lambda x, \quad x \neq 0$$

Solving for eigenvectors : pdf 120

1. determine λ's.
2. solve for X for each eigenvalue.
3. make sure the vectors have unit length. If not, divide the vector with $\sqrt{L_x}$.

the length of the corresponding eigenvector

Quadratic Form :

$Q(x) = x'Ax$, x is a $1 \times k$ vector and A is $k \times k$ matrix

$= \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_i x_j$

Any symmetric square matrix can be reconstructured from its eigenvalues and eigenvectors. The particular expression reveals the relative importance of each pair according to the relative size of the eigenvalue and the direction of the eigenvector.

Spectral Decomposition

$A = \sum_{i=1}^{k} \lambda_i e_i e_i'$, λ_i = eigenvalues , e_i = corresponding eigenvector

The ideas that lead to the spectral decomposition can be extended to provide a decomposition for a rectangular, rather than a square, matrix. If **A** is a rectangular matrix, then the vectors in the expansion of **A** are the eigenvectors of the square matrices **AA'** and **A'A**.

Singular-Value Decomposition :

- Let A be $m \times k$ rectangular matrix, then there exists an $m \times m$ orthogonal matrix U and $k \times k$ orthogonal matrix V such that,

$A = U \Lambda V'$, Λ = $m \times k$ diagonal matrix with λ_i's.

the positive constants λ_i are called the singular values of A.

- Assume there are r positive constants $\lambda_1, \lambda_2, \ldots, \lambda_r$, and r orthogonal $m \times 1$ unit vectors $u_1, u_2, \ldots, u_r$, and r orthogonal $k \times 1$ unit vectors $v_1, v_2, \ldots, v_r$,

$A = \sum_{i=1}^{r} \lambda_i u_i v_i' = U_r \Lambda_r V_r'$, here AA' has eigenvalue-eigenvector pairs (λ_i^2, u_i) that,

$AA' u_i = \lambda_i^2 u_i$, and $v_i = \lambda_i^{-1} A' u_i$.

(read PDF 122 for further details)

1. Get u_i by solving for eigenvalues $\gamma_i = \lambda_i^2$ of AA'

2. Get v_i by solving for eigenvalues $\gamma_i = \lambda_i^2$ of $A'A$

Then A would be,

$A = \sum \lambda_i u_i v_i$

The equality may be checked by carrying out the operations on the right-hand side.

The singular-value decomposition is closely connected to a result concerning the approximation of a rectangular matrix by a lower-dimensional matrix, due to Eckart and Young ([2]). If a $m \times k$ matrix $\mathbf{A}$ is approximated by $\mathbf{B}$, having the same dimension but lower rank, the sum of squared differences

$$\sum_{i=1}^{m} \sum_{j=1}^{k} (a_{ij} - b_{ij})^2 = \text{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})']$$

Result 2A.16. Let $\mathbf{A}$ be an $m \times k$ matrix of real numbers with $m \geq k$ and singular value decomposition $\mathbf{U\Lambda V}'$. Let $s < k = \text{rank}(\mathbf{A})$. Then

$$\mathbf{B} = \sum_{i=1}^{s} \lambda_i \mathbf{u}_i \mathbf{v}_i'$$

is the rank-s least squares approximation to $\mathbf{A}$. It minimizes

$$\text{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})']$$

over all $m \times k$ matrices $\mathbf{B}$ having rank no greater than s. The minimum value, or error of approximation, is $\sum_{i=s+1}^{k} \lambda_i^2$. ■

To establish this result, we use $\mathbf{UU}' = \mathbf{I}_m$ and $\mathbf{VV}' = \mathbf{I}_k$ to write the sum of squares as

$$\begin{aligned} \text{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})'] &= \text{tr}[\mathbf{UU}'(\mathbf{A} - \mathbf{B})\mathbf{VV}'(\mathbf{A} - \mathbf{B})'] \\ &= \text{tr}[\mathbf{U}'(\mathbf{A} - \mathbf{B})\mathbf{VV}'(\mathbf{A} - \mathbf{B})'\mathbf{U}] \\ &= \text{tr}[(\mathbf{\Lambda} - \mathbf{C})(\mathbf{\Lambda} - \mathbf{C})'] = \sum_{i=1}^{m} \sum_{j=1}^{k} (\lambda_{ij} - c_{ij})^2 = \sum_{i=1}^{m} (\lambda_i - c_{ii})^2 + \sum_{i \neq j}\sum c_{ij}^2 \end{aligned}$$

where $\mathbf{C} = \mathbf{U}'\mathbf{BV}$. Clearly, the minimum occurs when $c_{ij} = 0$ for $i \neq j$ and $c_{ii} = \lambda_i$ for the s largest singular values. The other $c_{ii} = 0$. That is, $\mathbf{UBV}' = \mathbf{\Lambda}_s$ or $\mathbf{B} = \sum_{i=1}^{s} \lambda_i \mathbf{u}_i \mathbf{v}_i'$.