Chapter 2. A Short Excursion into Matrix Algebra

Applied Multivariate Statistical Analysis

Course Instructor: Shuen-Lin Jeng

Department of Statistics
National Cheng Kung University

January, 10, 2021

2.1 Elementary Operations

A matrix $\mathcal{A}$ is a system of numbers with n rows and p columns:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \cdots & a_{22} & \cdots & \\ \vdots & \ddots & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

We also write (a_{ij}) for $\mathcal{A}$ and denoted as $x$or $x(p \times 1)$.

2.1 Elementary Operations

Matrix Operations

Elementary operations are summarized below :

$$\mathcal{A}^T = (a_{ji})$$

$$\mathcal{A} + \mathcal{B} = (a_{ij} + b_{ij})$$

$$\mathcal{A} - \mathcal{B} = (a_{ij} - b_{ij})$$

$$c \cdot \mathcal{A} = (c \cdot a_{ij})$$

$$\mathcal{A} \cdot \mathcal{B} = \mathcal{A}(n \times p)\mathcal{B}(n \times m) = (c_{ij}) = \left(\sum_{j=1}^{p} a_{ij} b_{jk} \right)$$

2.1 Elementary Operations

Properties of Matrix Operations

$$(\mathcal{A}^T)^T = \mathcal{A}$$

$$(\mathcal{A}\mathcal{B})^T = \mathcal{B}^T\mathcal{A}^T$$

2.1 Elementary Operations

Table 2.1 Special matrices and vectors

Name	Definition	Notation	Example
Scalar	$p = n = 1$	a	3
Column vector	$p = 1$	a	$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$
Row vector	$n = 1$	$a^\top$	$\begin{pmatrix} 1 & 3 \end{pmatrix}$
Vector of ones	$(\underbrace{1, \ldots, 1}_{n})^\top$	1_n	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
Vector of zeros	$(\underbrace{0, \ldots, 0}_{n})^\top$	0_n	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
Square matrix	$n = p$	$\mathcal{A}(p \times p)$	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$
Diagonal matrix	$a_{ij} = 0, i \neq j, n = p$	$\operatorname{diag}(a_{ii})$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
Identity matrix	$\operatorname{diag}(\underbrace{1, \ldots, 1}_{p})$	$\mathcal{I}_p$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Unit matrix	$a_{ij} = 1, n = p$	$1_n 1_n^\top$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
Symmetric matrix	$a_{ij} = a_{ji}$		$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$
Null matrix	$a_{ij} = 0$	0	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
Upper triangular matrix	$a_{ij} = 0, i < j$		$\begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$
Idempotent matrix	$\mathcal{A}\mathcal{A} = \mathcal{A}$		$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$
Orthogonal matrix	$\mathcal{A}^\top \mathcal{A} = \mathcal{I} = \mathcal{A}\mathcal{A}^\top$		$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$

Matrix Characteristics

Rank

The rank of a matrix $\mathcal{A}(n \times p)$ is defined as the maximum number of linearly independent rows (columns). A set of k rows a_j of $\mathcal{A}(n \times p)$ are said to be linearly independent if $\sum_{j=1}^{k} c_j a_j = 0_p$ implies $c_j = 0, \forall j$, where $c_1, \cdots, c_k$ are scalars.

Trace

$$tr(A) = \sum_{i=1}^{p} a_{ii}$$

2.1 Elementary Operations

Determinant

$$det(A) = |A| = \sum (-1)^{|\tau|} a_{1\tau(1)} \cdots a_{p\tau(p)}$$

the summation is over all permutations τ of $\{1, 2, \ldots, p\}$, and $|\tau| = 0$ if the permutation can be written as a product of an even number if transpositions and $|\tau| = 1$ otherwise.

$$\left|\mathcal{A}^T\right| = |\mathcal{A}|$$

$$|\mathcal{A}\mathcal{B}| = |\mathcal{A}| \cdot |\mathcal{A}|$$

$$|c\mathcal{A}| = c^n |\mathcal{A}|$$

2.1 Elementary Operations

Transpose

For $\mathcal{A}(n \times p)$ and $\mathcal{B}(p \times n)$

$$(\mathcal{A}^T)^T = \mathcal{A} \text{ and } (\mathcal{A}\mathcal{B})^T = \mathcal{B}^T \mathcal{A}^T$$

2.1 Elementary Operations

Inverse

If $|\mathcal{A}| \neq 0$ and $\mathcal{A}(p \times p)$, then the inverse $\mathcal{A}^{-1}$ exists:

$$\mathcal{A}\mathcal{A}^{-1} = \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}_p$$

the inverse of $\mathcal{A} = (a_{ij})$:

$$\mathcal{A}^{-1} = \frac{C}{|\mathcal{A}|}$$

where $\mathcal{C} = (c_{ij})$ is the adjoint matrix of $\mathcal{A}$.

The elements c_{ji} of $\mathcal{C}^T$ are the co-factors of $\mathcal{A}$:

$$c_{ji} = (-1)^{i+j} \begin{vmatrix} a_{11} & \ldots & a_{1(j-1)} & a_{1(j+1)} & \ldots & a_{1p} \\ \vdots & & & & & \\ a_{(i-1)1} & \ldots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \ldots & a_{(i-1)p} \\ a_{(i+1)1} & \ldots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \ldots & a_{(i+1)p} \\ \vdots & & & & & \\ a_{p1} & \ldots & a_{p(j-1)} & a_{p(j+1)} & \ldots & a_{pp} \end{vmatrix}.$$

The relationship between determinant and inverse of matrix $\mathcal{A}$ is $|\mathcal{A}^{-1}| = |\mathcal{A}|^{-1}$.

G-inverse

G-inverse(Generalized Inverse) A^{-} which satisfies the following :

$$\mathcal{A}\mathcal{A}^{-}\mathcal{A} = \mathcal{A}$$

Example 2.2 The generalized inverse can also be calculated for singular matrices.We have

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

which means that the generalized inverse of $\mathcal{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is $\mathcal{A}^{-} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ even though the inverse matrix of $\mathcal{A}$ does not exist in this case.

2.1 Elementary Operations

Eigenvalues, Eigenvectors

Consider a $(p \times p)$ matrix A. If there a scalar λ and a vector γ exists such as

$$\mathcal{A}\gamma = \lambda\gamma \tag{2.1}$$

then we call

$$\lambda \text{ an eigenvalue}$$

$$\gamma \text{ an eigenvector.}$$

It can be proven that an eigenvalue λ is a root of the p-th order polynomial $|\mathcal{A} - \lambda I_p| = 0$. Therefore are up to p eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$ of $\mathcal{A}$. Let $\mathbf{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_p)$.

2.1 Elementary Operations

Eigenvalues, Eigenvectors

The determinant $|\mathcal{A}|$ and the trace $tr(\mathcal{A})$ can be rewritten in terms of the eigenvalues:

$$|\mathcal{A}| = |\mathbf{\Lambda}| = \prod_{j=1}^{p} \lambda_j \tag{2.2}$$

$$tr(\mathcal{A}) = tr(\mathbf{\Lambda}) = \sum_{j=1}^{p} \lambda_j \tag{2.3}$$

An idempotent matrix $\mathcal{A}$ (see the definition in Table 2.1) can only have eigenvalues in $\{0, 1\}$; therefore, $tr(\mathcal{A}) = \text{rank}(\mathcal{A}) =$ number of eigenvalues $\neq 0$.

Eigenvalues, Eigenvectors

Example 2.3 Let us consider the matrix $\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. It is easy to verify that $\mathcal{A}\mathcal{A} = \mathcal{A}$ which implies that the matrix $\mathcal{A}$ is idempotent. We know that the eigenvalues of idempotent matrix are equal to 0 or 1.

In this case, the eigenvalues of $\mathcal{A}$ are $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 =$0 since

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = 1 \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \text{ and, } \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} = 0 \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Using formulas (2.2) and (2.3), we can calculate the trace and the determinant of A from the eigenvalues:

$\mathrm{tr}(\mathcal{A}) = \lambda_1 + \lambda_2 + \lambda_3 = 2$, $|\mathcal{A}| = \lambda_1\lambda_2\lambda_3 = 0$, and $\mathrm{rank}(\mathcal{A}) = 2$.

2.1 Elementary Operations

Properties of Matrix Characteristics

$\mathcal{A}(n \times n)$, $\mathcal{B}(n \times n)$, $\mathrm{c} \in \mathbb{R}$

$$tr(\mathcal{A} + \mathcal{B}) = tr(\mathcal{A}) + tr(\mathcal{B}) \tag{2.4}$$

$$tr(c\mathcal{A}) = c\, tr(\mathcal{A}) \tag{2.5}$$

$$|c\mathcal{A}| = c^n |\mathcal{A}| \tag{2.6}$$

$$|\mathcal{A}\mathcal{B}| = |\mathcal{B}\mathcal{A}| = |\mathcal{A}||\mathcal{B}| \tag{2.7}$$

$$rank(\mathcal{A}^T \mathcal{A}) = rank(\mathcal{A}) \tag{2.11}$$

$$rank(\mathcal{A} + \mathcal{B}) = rank(\mathcal{A}) + rank(\mathcal{B}) \tag{2.12}$$

$$rank(\mathcal{A}\mathcal{B}) = \min\{rank(\mathcal{A}), rank(\mathcal{B})\} \tag{2.13}$$

Properties of Matrix Characteristics

$\mathcal{A}(n \times p)$, $\mathcal{B}(p \times q)$, $\mathcal{C}(q \times n)$

$$tr(\mathcal{ABC}) = tr(\mathcal{BCA}) = tr(\mathcal{CAB}) \tag{2.14}$$

$$rank(\mathcal{ABC}) = rank(\mathcal{B}) \text{ for nonsingular } \mathcal{A}, \mathcal{C} \tag{2.15}$$

$\mathcal{A}(p \times p)$

$$|\mathcal{A}^{-1}| = |\mathcal{A}|^{-1} \tag{2.16}$$

$$rank(\mathcal{A}) = p \text{ if and only if } \mathcal{A} \text{ is nonsingular.} \tag{2.17}$$

2.2 Spectral Decompositions

Theorem 2.1 (Jordan Decomposition) *Each symmetric matrix $\mathcal{A}(p \times p)$ can be written as*

$$\mathcal{A} = \Gamma \mathbf{\Lambda} \Gamma^T = \sum_{j=1}^{p} \lambda_j \gamma_j \gamma_j^T \tag{2.18}$$

where

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_p)$$

and where

$$\Gamma = (\gamma_1, \gamma_2, \cdots, \gamma_p)$$

is an orthogonal matrix consisting of the eigenvectors γ_j of $\mathcal{A}$.

For some $\alpha \in \mathbb{R}$

$$\mathcal{A}^\alpha = \Gamma \mathbf{\Lambda}^\alpha \Gamma^T \tag{2.19}$$

where $\mathbf{\Lambda}^\alpha = \text{diag}(\lambda_1^\alpha, \cdots, \lambda_p^\alpha)$.

Theorem 2.2 (Singular Value Decomposition) *Each matrix $\mathcal{A}(n \times p)$ with rank r can be decomposed as*

$$\mathcal{A} = \Gamma \boldsymbol{\Lambda} \Delta^T,$$

where $\Gamma(n \times r)$ and $\Delta(p \times r)$. Both Γ and Δ are column orthogonal , i.e., $\Gamma^T\Gamma = \Delta^T\Delta = \mathcal{I}_r$ and $\boldsymbol{\Lambda} = diag(\lambda_1^{1/2}, \cdots, \lambda_r^{1/2}), \lambda_j > 0$. The values $\lambda_1, \cdots, \lambda_r$ are the nonzero eigenvalues of the matrices $\mathcal{A}\mathcal{A}^\mathcal{T}$ and $\mathcal{A}^\mathcal{T}\mathcal{A}$. Γ and Δ consists of the corresponding r eigenvectors of these matrices.

Properties of Matrix Characteristics

Example2.5 In Example 2.2, we showed that the generalized inverse of $\mathcal{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is $\mathcal{A}^{-} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

The following also holds $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ which means that the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 8 \end{pmatrix}$ is also a generalized inverse of $\mathcal{A}$.

2.3 Quadratic Forms

A quadratic form $\mathcal{Q}(x)$ is built form a symmetric matrix $\mathcal{A}(p \times p)$ and a vector $x \in \mathbb{R}^p$:

$$\mathcal{Q}(x) = x^T \mathcal{A} x = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} x_i x_j. \tag{2.21}$$

Definiteness of Quadratic Forms and Matrices

$$Q(x) > 0 \text{ for all } x \neq 0 \textit{ positive definite}$$

$$Q(x) \geq 0 \text{ for all } x \neq 0 \textit{ positive semidifinite}$$

A matrix $\mathcal{A}$ is called positive definite (semidefinite) if the corresponding quadratic form $\mathcal{Q}(\cdot)$ is positive definite (semidefinite).We write $\mathcal{A} > 0 (\geq 0)$.

Theorem 2.4 $\mathcal{A} > 0$ of and only if all $\lambda_i > 0$, $i = 1, \cdots, p$.

Corollary 2.1 If $\mathcal{A} > 0$, then $\mathcal{A}^{-1}$ exists and $|A| > 0$.

Theorem 2.5 *If $\mathcal{A}$ and $\mathcal{B}$ are symmetric and $\mathcal{B} > 0$, then the maximum of $\frac{x^T \mathcal{A} x}{x^T \mathcal{B} x}$ is given by the largest eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$. More generally.*

$$max_x \frac{x^T \mathcal{A} x}{x^T \mathcal{B} x} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = min_x \frac{x^T \mathcal{A} x}{x^T \mathcal{B} x}$$

where $\lambda_1, \ldots \lambda_p$ denote the eigenvalues of $\mathcal{B}^{-1}\mathcal{A}$.The vector which maximizes(minimizes) $\frac{x^T \mathcal{A} x}{x^T \mathcal{B} x}$ is the eigenvector of $\mathcal{B}^{-1}\mathcal{A}$ which corresponds to the largest (smallest) eigenvalue of $\mathcal{B}^{-1}\mathcal{A}$. If $x^T \mathcal{B} x = 1$,we get

$$max_x x^T \mathcal{A} x = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = min_x x^T \mathcal{A} x$$

2.4 Derivatives

The Hessian of the quadratic form $\mathcal{Q}(x) = x^T \mathcal{A} x$ is

$$\frac{\partial^2 x^T \mathcal{A} x}{\partial x \partial x^T} = 2\mathcal{A}. \tag{2.25}$$

2.5 Partitioned Matrices

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix},$$

where $\mathcal{A}_{ij}(n_i \times p_j)$, i,j =1,2 ,$n_1 + n_2 = n$ and $p_1 + p_2 = p$.

$$\mathcal{A}^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix} \tag{2.26}$$

where

$$\begin{cases} \mathcal{A}^{11} = (\mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21})^{-1} \stackrel{\text{def}}{=} (\mathcal{A}_{11\cdot 2})^{-1} \\ \mathcal{A}^{12} = -(\mathcal{A}_{11\cdot 2})^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1} \\ \mathcal{A}^{21} = -\mathcal{A}_{22}^{-1}\mathcal{A}_{21}(\mathcal{A}_{11\cdot 2})^{-1} \\ \mathcal{A}^{22} = \mathcal{A}_{22}^{-1} + \mathcal{A}_{22}^{-1}\mathcal{A}_{21}(\mathcal{A}_{11\cdot 2})^{-1}\mathcal{A}_{12}\mathcal{A}_{22}^{-1} \end{cases}$$

2.5 Partitioned Matrices

The following results will be useful if A_{11} is nonsingular:

$$|\mathcal{A}| = |\mathcal{A}_{11}| \left|\mathcal{A}_{22} - \mathcal{A}_{21}\mathcal{A}_{11}^{-1}\mathcal{A}_{12}\right| = |\mathcal{A}_{11}| \, |\mathcal{A}_{22\cdot1}| \tag{2.27}$$

If A_{22} is nonsingular, we have that

$$|\mathcal{A}| = |\mathcal{A}_{22}| \left|\mathcal{A}_{11} - \mathcal{A}_{12}\mathcal{A}_{22}^{-1}\mathcal{A}_{21}\right| = |\mathcal{A}_{22}| \, |\mathcal{A}_{11\cdot2}| \tag{2.28}$$

2.6 Geometrical Aspects

Distance

Let x,y $\in \mathbb{R}^p$. A distance d is defined as a function

$$d : \mathbb{R}^{2p} \to \mathbb{R}_+ \quad \text{which fulfills} \quad \begin{cases} d(x,y) > 0 & \forall x \neq y \\ d(x,y) = 0 & \text{if and only if } x = y \\ d(x,y) \leq d(x,z) + d(z,y) & \forall x, y, z \end{cases}$$

A *Euclidean distance d* between two points x and y is defined as

$$d^2(x,y) = (x-y)^T \mathcal{A}(x-y) \tag{2.32}$$

where $\mathcal{A}$ is a positive-definite matrix ($\mathcal{A}>0$). A is called a *metric*.

Example 2.10 A particular case is when $\mathcal{A} = \mathcal{I}_p$, i.e.,

$$d^2(x,y) = \sum_{i=1}^{p} (x_i - y_i)^2 \tag{2.33}$$

2.6 Geometrical Aspects

Fig. 2.1 Distance d

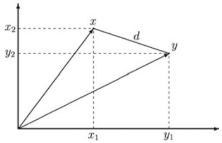
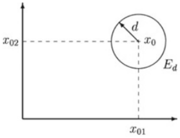


Fig. 2.2 Iso-distance sphere



A positive-definite matrix $A(A > 0)$leads to the iso-distance curves

$$E_d = \left\{x \in \mathbb{R}^p \mid (x - x_0)^T \mathcal{A}(x - x_0) = d^2\right\} \tag{2.34}$$

i.e., ellipsoids with center x_0, matrix A and constant d.

2.6 Geometrical Aspects

Theorem 2.7

(i) The principal axes of E_d *are in the direction of* γ_i ; *i* = *1,...,p.*

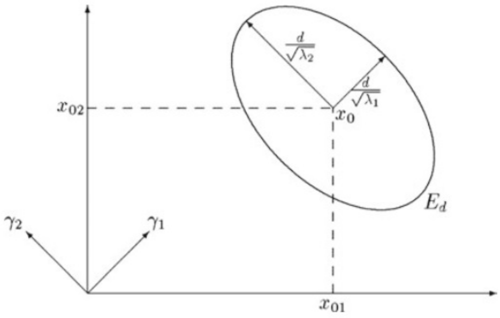
(ii) The half-lengths of the axes are; $\sqrt{\frac{d^2}{\lambda_i}}$ *i* = *1,...,p.*

(iii) The rectangle surrounding the ellipsoid E_d *is defined by the following inequalities:*

$$x_{0i} - \sqrt{d^2 a^{ii}} \leq x_i \leq x_{0i} + \sqrt{d^2 a^{ii}}, i = 1, \ldots, p,$$

where a^{ii} *is the (i,i) element of* A^{-1}. *By the rectangle surrounding the ellipsoid* E_d *we mean the rectangle whose side are parallel to the coordinate axis.*

Fig. 2.3 Iso–distance ellipsoid



2.6 Geometrical Aspects

Norm of a Vector

The norm or length of x (with respect to the metric $\mathcal{I}_p$) is defined as

$$\|x\| = d(0_p, x) = \sqrt{x^T x}$$

If $\|x\| = 1$, x is called a *unit vector* .A more general norm can be defined with respect to the metric $\mathcal{A}$:

$$\|x\|_{\mathcal{A}} = \sqrt{x^T \mathcal{A} x}$$

Angle between two Vectors

$$\cos\theta = \frac{x^T y}{\|x\| \, \|y\|}, \tag{2.40}$$

The angle can also be defined with respect to a general metric $\mathcal{A}$

$$\cos\theta = \frac{x^T \mathcal{A} y}{\|x\|_{\mathcal{A}} \, \|y\|_{\mathcal{A}}} \tag{2.43}$$

2.6 Geometrical Aspects

Example 2.11 Assume that these are two centered (i.e.,zero mean) data vectors. The cosine of the angle between them is equal to their correlation (defined in (3.8)). Indeed for x and y with $\bar{x} = \bar{y} = 0$ we have

$$r_{XY} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \cos\theta$$

Rotations

Let Γ be a (2×2) orthogonal matrix where

$$\Gamma = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{2.44}$$

If the axes are rotated about the origin through an angle θ in a clockwise direction, the new coordinates of P will be given by the vector y

$$y = \Gamma x \tag{2.45}$$

and a rotation through the same angle in a anticlockwise direction gives the new coordinates as

$$y = \Gamma^T x \tag{2.46}$$

Column Space and Null Space of a Matrix

Define of $\mathcal{X}$ $(n \times p)$

$$\mathrm{Im}(\mathcal{X}) \overset{def}{=} C(\mathcal{X}) = \{x \in \mathbb{R}^n \mid \exists a \in \mathbb{R}^p \text{ so that } Xa = x\}.$$

the space generated by the columns of $\mathcal{X}$ or the column space of $\mathcal{X}$. Note that $C(\mathcal{X}) \subseteq \mathbb{R}^n$ and $dim\{C(\mathcal{X})\} = rank(\mathcal{X}) = r \leq min(n, p)$.

$$\mathrm{Ker}(\mathcal{X}) \overset{def}{=} N(\mathcal{X}) = \{y \in \mathbb{R}^p \mid \mathcal{X}_y = 0\}$$

is the *null space* of $\mathcal{X}$.

Note that $N(\mathcal{X}) \subseteq \mathbb{R}^p$ and that $dim\{N(\mathcal{X})\} = p - r$.

2.6 Geometrical Aspects

Remark 2.2 $N(\mathcal{X}^T)$ is the orthogonal complement of $C(\mathcal{X})$ in $\mathbb{R}^n$, i.e.,given a vector b $\in \mathbb{R}^n$ it will hold that $x^T b = 0$ for all x $\in C(\mathcal{X})$, if and only if b $\in N(\mathcal{X}^T)$.

Example 2.12 Let $\mathcal{X} = \begin{pmatrix} 2 & 3 & 5 \\ 4 & 6 & 7 \\ 6 & 8 & 6 \\ 8 & 2 & 4 \end{pmatrix}$. It is easy to show (e.g.,by calculating the determinant of $\mathcal{X}$) that $rank(\mathcal{X}) = 3$. Hence, the column space of $\mathcal{X}$ is $C(\mathcal{X}) = \mathbb{R}^3$. The null space of $\mathcal{X}$ contains only the zero vector $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^T$ and its dimensions is equal to $rank(\mathcal{X})$ - 3 = 0.

2.6 Geometrical Aspects

Projection Matrix

A matrix $\mathcal{P}$ ($n \times n$) is called an (orthogonal) projection matrix in $\mathbb{R}^n$ if and only if $\mathcal{P} = \mathcal{P}^T = \mathcal{P}^2$ ($\mathcal{P}$ is idempotent). Let b $\in \mathbb{R}^n$. Then a $= \mathcal{P}b$ is the projection of b on $C(\mathcal{P})$.

2.6 Geometrical Aspects

Projection on $C(\mathcal{X})$

Consider $\mathcal{X}$ $(n \times p)$ and let

$$\mathcal{P} = \mathcal{X}(\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T \tag{2.47}$$

and $\mathcal{Q} = \mathcal{I}_n$ - $\mathcal{P}$. It's easy to check that $\mathcal{P}$ and $\mathcal{Q}$ are idempotent and that

$$\mathcal{P}\mathcal{X} = \mathcal{X} \text{ and } \mathcal{Q}\mathcal{X} = 0. \tag{2.48}$$

Since the columns of $\mathcal{X}$ are projected onto themselves, the projection matrix $\mathcal{P}$ projects any vector $\mathrm{b} \in \mathbb{R}^n$ onto $C(\mathcal{X})$. Similarly, the projection matrix $\mathcal{Q}$ projects any vector $\mathrm{b} \in \mathbb{R}^n$ onto the orthogonal complement of $C(\mathcal{X})$.

2.6 Geometrical Aspects

Theorem 2.8 *Let $\mathcal{P}$ be the Projection (2.47) and $\mathcal{Q}$ its orthogonal complement. Then:*

(i) x $= \mathcal{P}b$ entails x $\in C(\mathcal{X})$,

(ii) y $= \mathcal{Q}b$ means that $y^T x = 0$ $\forall$x $\in C(\mathcal{X})$.

Remark 2.3 Let x,y $\in \mathbb{R}^n$ and consider $p_x \in \mathbb{R}^n$, the projection of x on y (see Fig.2.5). With $\mathcal{X} =$ y we have form(2.47)

$$p_x = y(y^T y)^{-1} y^T x = \frac{y^T x}{\|y\|^2} y \tag{2.49}$$

and we can easily verify that

$$\|p_x\| = \sqrt{p_x^T p_x} = \frac{|y^T x|}{\|y\|}.$$