MOORE-PENROSE GENERALIZED INVERSES
AND
APPLICATIONS

by

Amani A. N. Al_Akhdar

Supervisor: Dr. Taha Abu_kaff

Submitted as a partial fulfillment of the

requirements for the degree of

Master Of Science

in

Mathematics

at

Department of Mathematics
College Of Science & Technology
Al_Quds University
Abu Dies_Jerusalem

Jerusalem, 4/8/2001.

Abstract

We studied a special kind of generalized inverses of an $m \times n$ matrix on a real or a complex field, which is the Moore-Penrose inverses.

The existence, uniqueness and the basic properties of such inverses also studied.

We discussed the generalized inverses and their relation in solving linear systems.

Furthermore, we discussed some methods for finding the Moore-Penrose inverse of any matrix.

Finally, we studied two fields of applications of the Moore-Penrose inverses. The first finds minimum least-squares solution of a linear system, and the second in statistics.

Contents

CHAPTER ONE

Preliminaries

This chapter contains definitions theorems and ideas that we shall need in the following chapters; which consists of three sections, some basic definitions and theorems are listed in the first section, section two about singular value decomposition of a matrix. One sided inverses discussed in the third section.

(1.1) Basic definitions and theorems :

Definition (1.1.1) :

A square matrix A is said to be *Hermitian* if and only if $A = A^*$, where $A^* = (\overline{A})^T$ (the conjugate transpose).

Definition (1.1.2) :

A square matrix A that satisfy the equation $A^2 = A$, is said to be *Idempotent* matrix, or (*Projector*).

Theorem (1.1.1) :

If P is an idempotent matrix, then :

(a) $I - P$ is idempotent;

(b) $\mathrm{Im}(I - P) = KerP$;

(c) $Ker(I - P) = \mathrm{Im}\,P$;

Proof : (a) since $(I - P)^2 = I - 2P + P^2 = I - 2P + P = I - P$, so $I - P$ is idempotent.

(b) if $y \in \mathrm{Im}(\mathrm{I}-P)$ then there exist $x \in F^n$ such that $(\mathrm{I}-P)x = y$. Hence $Py = P(\mathrm{I}-P)x = (P-P^2)x = 0$, so $y \in KerP$. Conversely if $Py = 0$ then $(\mathrm{I}-P)y = y$, so $y \in \mathrm{Im}(\mathrm{I}-P)$, Thus $\mathrm{Im}(\mathrm{I}-P) = KerP$.

(c) This part can be proved in a similar way as part (b).

Definition (1.1.3) :

For an idempotent matrix P, the idempotent matrix $Q = \mathrm{I}-P$ is called the *complementary projector* to P.

Theorem (1.1.2) :

If P is a projector, then

$$KerP \dotplus \mathrm{Im}\, P = F^n$$

where F^n is the space on which P acts

Proof : Let $x \in F^n$ then $x = x_1 + x_2$ where

$$x_1 = (\mathrm{I}-P)x \in KerP \qquad \text{and} \qquad x_2 = Px \in \mathrm{Im}\, P.$$

on the other hand if $x \in KerP + \mathrm{Im}\, P$ then $x \in F^n$, hence $F^n = KerP + \mathrm{Im}\, P$. Now we need to show that the sum is direct, if $x \in KerP \cap \mathrm{Im}\, P$ then $x \in KerP$ or $Px = 0$, and $x \in \mathrm{Im}\, P$ then $x \in Ker(\mathrm{I}-P)$ or $(\mathrm{I}-P)x = 0$, hence by these two conclusions we have

$$Px + (\mathrm{I}-P)x = 0 \qquad \text{or} \qquad x = 0$$

and so the sum is direct, or $F^n = KerP \dotplus \mathrm{Im}\, P$. $\square$

Note : We can say that the idempotent matrix P performs a projection of the space F^n onto $\mathrm{Im}\, P$ along $KerP$.

Definition (1.1.4) :

A matrix $P \in F^{n \times n}$ is said to be *orthogonal projector* (or *projection*)if it is hermitian idempotent matrix.

Definition (1.1.5) :

A matrix $U \in F^{n \times n}$ is said to be *unitary* or *orthogonal* if $U^*U = \mathrm{I}$ and $UU^* = \mathrm{I}$.

Definition (1.1.6):

A symmetric matrix A is *positive definite* if for every nonzero vector $\mathbf{x}$,

$$\mathbf{x}^* A\mathbf{x} > 0$$

and is *positive semidefinite* if $\mathbf{x}^* A\mathbf{x} \geq 0$.

Theorem (1.1.3) :

If A and B are two matrices such that ahe matrix product AB are defined, then

$$\operatorname{rank}(\mathrm{AB}) \leqslant \min(\operatorname{rank}\mathrm{A}, \operatorname{rank}\mathrm{B}).$$

Lemma (1.1.1) : [1]

If $\mathrm{A} \in F^{m \times n}$ then

$$\operatorname{rank}\mathrm{AA}^* = \operatorname{rank}\mathrm{A} = \operatorname{rank}\mathrm{A}^*\mathrm{A}.$$

Definition (1.1.7) :

Let $\mathcal{L}$ be a finite dimensional linear space over the field F and let $\mathbf{x}, \mathbf{y} \in \mathcal{L}$. A binary operation $\langle \mathbf{x}, \mathbf{y} \rangle$ from $\mathcal{L} \times \mathcal{L}$ to F is said to be an inner product on $\mathcal{L}$ if the following properties are satisfied for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{L}$, and $\alpha, \beta \in F$:

(a) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$,

(b) $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$,

(c) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$, where * denotes the complex conjucat.

Theorem (1.1.4) :

If $\mathrm{A} \in F^{n\times n}$, then $\langle x, \mathrm{A}y \rangle = \langle \mathrm{A}^* x, y \rangle$ for all $x, y \in F^n$.

Definition (1.1.8) :

Let $\mathbf{x}$ be an $n\times 1$ vector then the *Euclidean norm* of $\mathbf{x}$ is

$$\|\mathbf{x}\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2}$$

Definition (1.1.9) :

A nonzero square matrix P is called a *permutation matrix* if there is exactly one nonzero entry in each row and column that is 1 and if the rest are all zeros.

Definition (1.1.10) :

A matrix in $F_r^{m\times n}$ is said to be in *Hermite normal from* (or *reduced row echelon from*) if:

(i) Each of the first r rows contains at least one nonzero element ; the remaining rows contain only zeros.

(ii) The first r columns of the identity matrix I_m appear in the first r columns.

Theorem (1.1.5) :

For any $m\times n$ matrix A there exist a nonsingular $m\times m$ matrix P and a nonsingular $n\times n$ matrix V such that PAV is one of the matrices

$$[\mathrm{I}_m \quad 0], \quad \begin{bmatrix} \mathrm{I}_n \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{bmatrix} \text{(where } r = \operatorname{rank} \mathrm{A}\text{), and } \quad \mathrm{I}_n \text{ (if } m = n\text{).}$$

(1.2) Singular value decomposition

Theorem (1.2.1) :

The eigenvalues of the matrix A^*A are real and nonnegative.

Definition (1.2.1) :

Let A be $m\times n$ matrix, $(m \geq n)$, then the eigenvalues of the $n\times n$ symmetric matrix A^*A are denoted by $\sigma_i^{\,2}$, $i = 1,2,\ldots,n$, where $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2$. Then $\sigma_1, \sigma_2, \ldots, \sigma_n$ are called the *singular values* of A.

Theorem (1.2.2) :[13]

Let A be $m\times n$ matrix and let $\{\sigma_i\}_{i=1}^{r}$ be the nonzero singular values of A. Then A can be represented in the form

$$A = U\Sigma V^*,$$

where $U \in F^{m\times m}$ and $V \in F^{n\times n}$ are unitary and the $m\times n$ matrix Σ has σ_i in the i,i position $(1 \leq i \leq r)$ and zeros elsewhere.

The representation $A = U\Sigma V^*$ is denoted as the *singular value decomposition* of A.

Note : The columns of the matrix V are orthonormal eigenbasis of A^*A, and the columns of U are orthonormal eigenbasis of AA^* that can be obtained by solving the eigenvalue-eigenvector problem for the matrices A^*A and AA^*.

(1.3) One sided inverse

In this section we shall introduce a generalization of the inverse matrix not only for nonsingular (square) matrices but for singular square and rectangular matrices as well; which is the notion of one sided invertibility of matrices.

Definition (1.3.1) :

A matrix $A \in F^{m \times n}$ is said to be *left* (respectively, *right*) *Invertible* if there exists a matrix A_L^{-1} (respectively, A_R^{-1}) from $F^{n \times m}$ such that

$$A_L^{-1}A = I_n \text{ (respectively , } AA_R^{-1} = I_m \text{)} \qquad (1.3.1)$$

A matrix A_L^{-1} (respectively, A_R^{-1}) satisfying equation (1.3.1) is called a *left* (respectively, *right*) *inverse* of A. If $m = n$ and A is a nonsingular matrix then

$$A_L^{-1} = A_R^{-1} = A^{-1}$$

Theorem (1.3.1) :

Let $A \in F^{m \times n}$, then the following statements are equivalent:

(a) the matrix A is left invertible;

(b) $m \geq n$ and rank $A = n$;

(c) the columns of A are linearly independent as members of F^m;

(d) $Ker A = \{0\}$.

Theorem (1.3.2): $A \in F^{m \times n}$, then the following statements are equivalent:

(a) the matrix A is right invertible;

(b) $m \leq n$ and rank $A = m$.

(c) The rows of A are linearly independent as members of F^n;

(d) $\operatorname{Im} A \in F^m$.

CHAPTER FOUR

Applications of Moore-Penrose inverse

(4.1) Least-Square Problem

Consider the system $A\mathbf{x} = \mathbf{b}$, where the matrix A is nonsquare or singular. In such cases, solutions may not exist at all; in cases where there are solutions, there may be infinitely many.

In these cases, the best we can hope for is to find a vector $\mathbf{x}$ that will make $A\mathbf{x}$ as close as possible to the vector $\mathbf{b}$. In other words, we seek a vector $\mathbf{x}$ such that $\|r(\mathbf{x})\|^2 = \|A\mathbf{x} - \mathbf{b}\|^2$ is minimized, which is the sum of squares, and it is clear that minimizing $\|A\mathbf{x} - \mathbf{b}\|^2$ is equivalent to minimizing $\|A\mathbf{x} - \mathbf{b}\|$. When the Euclidean norm $\|\,.\,\|_2$ is used, this solution is said to be the *Least-squares solution* to the system $A\mathbf{x} = \mathbf{b}$. The problem of finding least-squares solutions to the linear system $A\mathbf{x} = \mathbf{b}$ is known as the *linear least-squares problem.*

Theorem(4.1.1):[3], Least-squares Existence and Uniqueness theorem

If A is an $m \times n$ matrix $(m > n)$ there always exists a solution to the linear least-square problem. This solution is unique if and only if A has full rank. If A is rank deficient (or $\operatorname{rank} A < n$)then the least-square problem has infinitely many solutions.

Remark: The solutions mentioned in the previous theorem obtained using the singular value decomposition of $A = U\Sigma V^*$, by solving the systems

$\mathbf{b}' = U^*\mathbf{b}$ and

$$y_i = \begin{cases} \dfrac{b_i'}{\sigma_i}, & if\ \sigma_i \neq 0 \\ arbitrary, & if\ \sigma_i = 0 \end{cases} \quad \text{and}$$

$\mathbf{x} = V\mathbf{y}$,

where $\mathbf{b}' = [b_1' \ \ b_2' \ \ \cdots \ \ b_m']^T$ and $\mathbf{y} = [y_1 \ \ y_2 \ \ \cdots \ \ y_n]^T$.

Lemma (4.1.1): [3]

$\mathbf{x}$ is a least-squares solution of the system $\mathrm{A}\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x}$ satisfies

$$\mathrm{A}^*\mathrm{A}\mathbf{x} = \mathrm{A}^*\mathbf{b}$$

minimum Least-Square solution

Theorem(4.1.1):

Let A be an $m \times n$ matrix, if A is of full rank, then the unique least-square solution is given by $\mathbf{x} = \mathrm{A}^+\mathbf{b}$. Else, there is a unique element $\mathbf{x}$ among the infinite solution set of minimal euclidean norm and this element is given by $\mathbf{x} = \mathrm{A}^+\mathbf{b}$.

Proof: This can be proved by showing that $\mathbf{x} = \mathrm{A}^+\mathbf{b}$ is a solution of the least-square problem, or that the given $\mathbf{x}$ satisfies $\mathrm{A}^*\mathrm{A}\mathbf{x} = \mathrm{A}^*\mathbf{b}$. Let $\mathrm{A} = U\Sigma V^*$ be the singular value decomposition of A, then the Moore-Penrose inverse $\mathrm{A}^+ = V\Sigma^+U^*$ (as mentioned in chapter three); now since $\Sigma^*\Sigma\Sigma^+ = \Sigma^*$, we find that

$$
\begin{aligned}
A^*Ax &= A^*AA^+\mathbf{b} \\
&= V\Sigma^*U^*\,U\Sigma V^*\,V\Sigma^+U^*\mathbf{b} \\
&= V\Sigma^*\,\Sigma\,\Sigma^+U^*\mathbf{b} \\
&= V\Sigma^*U^*\mathbf{b} \\
&= A^*\mathbf{b}
\end{aligned}
$$

This shows that $\mathbf{x} = A^+\mathbf{b}$ is a solution of the least-square problem (by Lemma (4.1.3)) .

To prove the next part. Let rank $A = r < n$. Using the same notation as earlier. we have seen that The least-square problem have an infinite set of solut-ions, we can get the solution $\mathbf{x}$ of minimal Euclidean norm, from $\mathbf{y}$ with minimal norm where $\mathbf{x} = V\mathbf{y}$, $\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]^T$, and

$$
y_i = \begin{cases} \dfrac{b_i'}{\sigma_i}, & if\ \sigma_i \neq 0 \\ arbitrary, & if\ \sigma_i = 0 \end{cases}
$$

such $\mathbf{y}$ obtained by putting $y_i = 0$ for each $i = r+1, r+2, \dots, n$. This cho-ice of $\mathbf{y}$ can be written as $\mathbf{y} = \Sigma^+\mathbf{b}'$. The minimal norm solution is given by

$$
\mathbf{x} = V\mathbf{y} = V\Sigma^+\mathbf{b}' = V\Sigma^+U^*\mathbf{b} = A^+\mathbf{b} \qquad \square
$$

Example(4.1.1):

Solve the linear least-square problem to the system $A\mathbf{x} = \mathbf{b}$ using the singular value decomposition where

$$A = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 4 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix},$$

and find the minimal least-square solution .

Solution: first we find the singular value decomposition for the matrix A, the square root of the eigenvalues of the symmetric matrix A^*A are $\sqrt{30}$, so

$$\Sigma = \begin{bmatrix} \sqrt{30} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and by finding the orthonormal eigenbasis of A^*A and the orthonormal eigenbasis of AA^* we get that

$$U = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{5} \\ -1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{5} \\ 2/\sqrt{6} & 0 & -1/\sqrt{5} \end{bmatrix} \quad , \quad V = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}.$$

Next we solve the linear problem

$$\mathbf{b}' = U^*\mathbf{b} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & -1/\sqrt{3} & -1/\sqrt{53} \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix} = \begin{bmatrix} 24/\sqrt{6} \\ 0 \\ 0 \end{bmatrix}$$

now $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ where y_i is given by (4.2.1), so

$$y_1 = \frac{24}{\sqrt{6}} \div \sqrt{30} = \frac{4}{\sqrt{5}} \quad \text{and}$$

$y_2 = \gamma$ where γ is an arbitrary constant because $\sigma_2 = 0$ .thus

$$\mathbf{y} = \begin{bmatrix} 4/\sqrt{5} \\ \gamma \end{bmatrix}$$

finally

$$\mathbf{x} = V\mathbf{y} = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 4/\sqrt{5} \\ \gamma \end{bmatrix}$$

$$= \begin{bmatrix} (8+\gamma\sqrt{5})/5 \\ (-4+\gamma 2\sqrt{5})/5 \end{bmatrix}$$

which is the family of all solutions of the given linear system.

The minimal least-square solution is

$$\mathbf{x} = \mathrm{A}^{+}\mathbf{b}$$

$$= V\Sigma^{+}U^{*}\mathbf{b}$$

$$= \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{30} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & -1/\sqrt{3} & -1/\sqrt{53} \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} 1/15 & -1/15 & 2/15 \\ -1/30 & 1/30 & -1/15 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} 8/5 \\ -4/5 \end{bmatrix}.$$

(4.2) *Moore-Penrose inverses in statistics :*

In this section we'll discuss some of the statistical applications for the Moore-Penrose inverses .

Random vectors and some related statistical concepts:

We review some of the basic definitions and results in distribution theory which will be needed later.

Definition (4.2.1):

Let $x_1, \dots, x_n$ be n independent random variables, then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

is called *random vector* (i.e a vector-valued function on a probability space).

Definition (4.2.2) :

The vector of expected values of the x_i's is the *mean* vector of $\mathbf{x}$, denoted by $\boldsymbol{\mu}_x$; that is,

$$\boldsymbol{\mu}_x = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = E(\mathbf{x}) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_n) \end{bmatrix}.$$

Definition (4.2.3) :

A measure of the linear relationship between x_i and x_j is given by the *covariance* of x_i and x_j, which is denoted by $\operatorname{cov}(x_i, x_j) = \sigma_{ij}$ and is defined by

$$\sigma_{ij} = \operatorname{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)]. \qquad (4.2.1)$$

Definition(4.2.4) :

The matrix Ω_x which has σ_{ij} as the (i,j)th element, is called the *variance-covariance matrix* of $\mathbf{x}$ which is denoted by $\mathrm{var}(\mathbf{x})$ or $\mathrm{cov}(\mathbf{x},\mathbf{x})$ such that

$$\Omega_x = \mathrm{var}(\mathbf{x}) = E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^*].$$

Note that this matrix Ω_x is symmetric because $\sigma_{ij} = \sigma_{ji}$ (by (4.2.1)), and Ω is diagonal if the x_i's are independent.

Suppose now that $\mathbf{y}$ is an m-dimensional random vector defined by a linear transformation on $\mathbf{x}$:

$$\mathbf{y} = \mathrm{A}\mathbf{x}$$

where A is an $m\times n$ matrix of constants ($m \le n$), then

$$\boldsymbol{\mu}_y = E(\mathbf{y}) = E(\mathrm{A}\mathbf{x}) = \mathrm{A}E(\mathbf{x}) = \mathrm{A}\boldsymbol{\mu}_x. \tag{4.2.2}$$

$$\begin{aligned}\Omega_y = \mathrm{var}(\mathbf{y}) &= E[(\mathbf{y}-\boldsymbol{\mu}_y)(\mathbf{y}-\boldsymbol{\mu}_y)^*] = E[(\mathrm{A}\mathbf{x}-\boldsymbol{\mu}_y)(\mathrm{A}\mathbf{x}-\boldsymbol{\mu}_y)^*] \\ &= E[(\mathrm{A}\mathbf{x}-\mathrm{A}\boldsymbol{\mu}_x)(\mathrm{A}\mathbf{x}-\mathrm{A}\boldsymbol{\mu}_x)^*] = E[\mathrm{A}(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^*\mathrm{A}^*] \\ &= \mathrm{A}E[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^*]\mathrm{A}^* = \mathrm{A}\Omega_x\mathrm{A}^*. \end{aligned} \tag{4.2.3}$$

The covariance matrix is positive semidefinite, that it is symmetric and since the variance is always nonnegative then for any vector of constants $\mathbf{t} = [t_1 \ \cdots \ t_n]^T$ we have

$$\mathbf{t}^*\Omega_x\mathbf{t} = \mathbf{t}^*E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^*]\mathbf{t} = E[\mathbf{t}^*(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^*\mathbf{t}] = \mathrm{var}(\mathbf{t}^*\mathbf{x})) \ge 0,$$

hence Ω_x is positive semidefinite .

The multivariate normal distribution :

We denote the univariate normal distribution by $N(\mu,,\sigma^2)$ where μ is the mean and σ^2 is the variance of the distribution, which may be zero. We define multivariate normal distribution as follows.

Definition (4.2.5) :

Let $z_1,\dots,z_k$ be independent $N(0,1)$ variables and define the vector variable

$$\mathbf{x} = \boldsymbol{\mu}_x + T\mathbf{z} \tag{4.2.4}$$

where $\mathbf{x}$ and $\boldsymbol{\mu}_x$ are the p-vectors, T is $p\times k$ matrix and $\mathbf{z}^* = [z_1 \ \cdots \ z_k]$, the distribution of $\mathbf{x}$ defined by (4.2.4) is called the p-variate normal with mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\Omega_x = TT^*$ which is represented by $N_p(\boldsymbol{\mu}_x,\Omega_x)$.

Theorem (4.2.1) :

Let the $p\times1$ random vector $\mathbf{x}$ have a p-variate normal distribution of rank k with mean $\boldsymbol{\mu}_x$ and covariance Ω_x. There exists a $k\times p$ matrix B and a $k\times1$ vector $\mathbf{b}$ such that the $k\times1$ vector $\mathbf{z}$ defined by $\mathbf{z} = B\mathbf{x}+\mathbf{b}$ has the standard k-variate normal distribution, that is, $\mathbf{z}$ is $N_k(\mathbf{0},\mathrm{I})$.

Proof: Since $\mathbf{x}$ has a normal distribution of rank k, then rank$\Omega_x = k$ and $\Omega_x = TT^*$ where T is a $p\times k$ matrix of rank k. Now define B to be T^+ and $\mathbf{b}$ to be $T^+\boldsymbol{\mu}_x$, then

$$\mathbf{z} = B\mathbf{x}+\mathbf{b} = T^+\mathbf{x} - T^+\boldsymbol{\mu}_x = T^+(\mathbf{x}-\boldsymbol{\mu}_x)$$

so

$$E(\mathbf{z}) = BE(\mathbf{x}) + \mathbf{b} = B\boldsymbol{\mu}_x + \mathbf{b} = T^{+}\boldsymbol{\mu}_x - T^{+}\boldsymbol{\mu}_x = \mathbf{0} \qquad (4.2.5)$$

and covariance matrix

$$\text{var}(\mathbf{z}) = B\Omega_x B^* = T^{+}\Omega_x (T^{+})^* = T^{+}TT^*(T^{+})^* = T^{+}T(T^{+}T)^*$$

but since T has a full rank, then $T^{+}T = \mathrm{I}$ (by theorem(3.2.3),(h)), and hence

$$\text{var}(\mathbf{z}) = \mathrm{I}, \qquad (4.2.6)$$

Using (4.2.5) and (4.2.6) we get that $\mathbf{z}$ is $N_k(\mathbf{0}, \mathrm{I})$. □

Definition (4.2.5) : The chi-square distribution

Let $x_1, \ldots, x_n$ be n independent normal variables with zero mean and unite variance (or $\text{var}(x_i) = 0$). Then the distribution of the statistic $\sum_{i=1}^{n} x_i^2$ is known as the central chi-square distribution with n degrees of freedom, and that of $\sum_{i=1}^{n} (x_i - v_i)^2$ is known as the noncentral chi-square distribution with n degrees of freedom and noncentrality parameter $\delta = \sum v_i^2$.

The Moore-Penrose inverse is useful in constructing quadratic forms, in normal random vectors, so that they have chi-squared distributions.

It is desired to determine whether or not the $m \times 1$ parameter vector $\boldsymbol{\mu}_t = \mathbf{0}$, for a sample statistic $\mathbf{t} \sim N_m(\boldsymbol{\mu}_t, \Omega_t)$, where Ω_t is positive definite, now let T be any $m \times m$ matrix satisfying $TT^* = \Omega_t$, and define $\mathbf{u} = T^{-1}\mathbf{t}$, then by (4.2.2)

$$E(\mathbf{u}) = E(T^{-1}\mathbf{t}) = T^{-1}E(\mathbf{t}) = T^{-1}\boldsymbol{\mu}_t$$

and by (4.2.3)

$$\text{var}(\mathbf{u}) = T^{-1}[\text{var}(\mathbf{t})](T^*)^{-1} = T^{-1}\Omega_t(T^*)^{-1} = T^{-1}(TT^*)(T^*)^{-1} = \mathrm{I}_m$$

so $\mathbf{u}$ has m-dimensional multivariate normal distribution with mean vector $T^{-1}\boldsymbol{\mu}_t$ and covariance matrix I_m or that $\mathbf{u} \sim \mathrm{N}_m(T^{-1}\boldsymbol{\mu}_t, \mathrm{I}_m)$. Consequantly, $u_1, u_2, \ldots, u_m$ are independently distributed random variables.

Consider the statistic

$$v_1 = \mathbf{t}^*\Omega_t^{-1}\mathbf{t}$$

and so

$$v_1 = \mathbf{t}^*\Omega_t^{-1}\mathbf{t} = \mathbf{t}^*(T^*)^{-1}T^{-1}\mathbf{t} = \mathbf{u}^*\mathbf{u} = \sum_{i=1}^{m} u_i^2$$

has a chi-squared distribution with m degrees of freedom. Which is central if $\boldsymbol{\mu}_t = \mathbf{0}$ and noncentral if $\boldsymbol{\mu}_t \neq \mathbf{0}$, but if Ω_t is positive semidefinite (may be singular), the construction of v_1 above can be generalized by using the Moore-Penrose inverse of Ω_t. Now let rank $\Omega_t = r$, then since Ω_t is symmetric then it's singular value decomposition is $\Omega_t = U\Sigma U^*$, we can write Ω_t as $\begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} = U_1\Sigma_1 U_1^*$, where U_1 is $m \times r$ matrix, U_2 is $m \times (n-r)$ matrix and Σ_1 is the nonsingular $r \times r$ matrix then $\Omega_t^+ = U_1\Sigma_1^{-1}U_1^*$. Define $\mathbf{w} = \Sigma_1^{-1/2}U_1^*\mathbf{t}$, since

$$E(\mathbf{w}) = E(\Sigma_1^{-1/2}U_1^*\mathbf{t}) = \Sigma_1^{-1/2}U_1^*E(\mathbf{t}) = \Sigma_1^{-1/2}U_1^*\boldsymbol{\mu}_t$$

and

$$\text{var}(\mathbf{w}) = \Sigma_1^{-1/2}U_1^*[\text{var}(\mathbf{t})]U_1\Sigma_1^{-1/2} = \Sigma_1^{-1/2}U_1^*(U_1\Sigma_1U_1^*)U_1\Sigma_1^{-1/2} = \mathrm{I}_r \ ,$$

then $\mathbf{w} \sim N_r(\Sigma_1^{-1/2}U_1^*\boldsymbol{\mu}_t, \mathrm{I}_r)$, thus since the $w_i s$ are independently distributed normal random variables,

$$v_2 = \mathbf{t}^*\Omega_t^+\mathbf{t} = \mathbf{t}^*U_1\Sigma_1^{-1}U_1^*\mathbf{t} = \mathbf{t}^*U\Sigma_1^{-1/2}\Sigma_1^{-1/2}U_1^*\mathbf{t} = \mathbf{w}^*\mathbf{w} = \sum_{i=1}^{r} w_i^2$$

has a chi-squared distribution, with r degrees of freedom, which is central if $\Sigma_1^{-1/2} U_1^* \boldsymbol{\mu}_t = \mathbf{0}$.