

# LINEAR MODEL NOTES

CARSON JAMES

## CONTENTS

1. Matrix Algebra	1
1.1. Column and Null Space	1
1.2. Generalized Inverses	3
1.3. Projections	4
1.4. Solving Linear Equations	7
1.5. Moore-Penrose Pseudoinverse	8
1.6. Differentiation	11
1.7. Quadratic Forms and Eigendecomposition	12
2. The Linear Model	13
2.1. Model Description	13
2.2. Least Squares Optimization	14
2.3. Estimation	15
2.4. Imposing Restrictions for a Unique Solution	18
2.5. Constrained Parameter Space	20
2.6. The Gauss-Markov Model	20
2.7. The Aitken Model	23
3. Distribution Theory	23
3.1. Introduction	23
3.2. Multivariate Normal	23
3.3. Chi-Square	25
3.4. Quadratic Forms	26

## 1. MATRIX ALGEBRA

### 1.1. Column and Null Space.

**Exercise 1.1.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $\mathcal{N}(X) = \mathcal{N}(X^T X)$ .

*Proof.* Let  $a \in \mathcal{N}(X)$ . Then  $Xa = 0$ . So  $X^T Xa = 0$ . Thus  $a \in \mathcal{N}(X^T X)$ . Conversely, suppose that  $a \in \mathcal{N}(X^T X)$ . Then  $X^T Xa = 0$ . So

$$\begin{aligned} 0 &= a^T X^T Xa \\ &= (Xa)^T (Xa) \\ &= \|Xa\|^2 \end{aligned}$$

Hence  $Xa = 0$  and  $a \in \mathcal{N}(X)$ . □

**Exercise 1.2.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $\mathcal{C}(X^T) = \mathcal{C}(X^T X)$ .

*Proof.*

$$\begin{aligned}\mathcal{C}(X^T) &= \mathcal{N}(X)^\perp \\ &= \mathcal{N}(X^T X)^\perp \\ &= \mathcal{C}(X^T X)\end{aligned}$$

□

**Exercise 1.3.** Let  $X \in \mathcal{M}_{m,n}$ . If  $X^T X = 0$ , then  $X = 0$ .

*Proof.* Suppose that  $X^T X = 0$ . Then

$$\begin{aligned}\text{rank}(X^T) &= \dim \mathcal{C}(X^T) \\ &= \dim \mathcal{C}(X^T X) \\ &= \text{rank}(X^T X) \\ &= 0\end{aligned}$$

So  $X^T = X = 0$ .

□

**Exercise 1.4.** Let  $X \in \mathcal{M}_{m,n}$  and  $A, B \in \mathcal{M}_{n,p}$ . Then  $X^T X A = X^T X B$  iff  $X A = X B$ .

*Proof.* Clearly if  $X A = X B$ , then  $X^T X A = X^T X B$ . Conversely, suppose that  $X^T X A = X^T X B$ . Then  $X^T X (A - B) = 0$ . So for each  $i = 1, \dots, p$ ,  $X^T X (A - B)e_i = 0$ . Thus for each  $i = 1, \dots, p$   $X(A - B)e_i \in \mathcal{N}(X^T) \cap \mathcal{C}(X) = \{0\}$ . Hence  $X(A - B) = 0$  and  $X A = X B$ . □

**Theorem 1.5.** Let  $X \in \mathcal{M}_{m,n}$ . Then

$$\text{nullity}(X) + \text{rank}(X) = n$$

.

**Exercise 1.6.** Let  $X \in \mathcal{M}_{m,n}$ . Then

$$\text{rank}(X^T) = \text{rank}(X)$$

*Proof.* We have that

$$\begin{aligned}\text{rank}(X^T) &= \text{rank}(X^T X) \\ &= n - \text{nullity}(X^T X) \\ &= n - \text{nullity}(X) \\ &= \text{rank}(X)\end{aligned}$$

□

**Definition 1.7.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $X$  is said to have **full column rank** if  $\text{rank}(X) = n$

**Exercise 1.8.** Let  $X \in \mathcal{M}_{m,n}$ . If  $X$  has full column rank, then

$$\mathcal{N}(X) = \{0\}$$

*Proof.* Suppose that  $X$  has full column rank. Then  $\text{rank}(X) = n$  Hence  $\text{nullity}(X) = 0$  and  $\mathcal{N}(X) = \{0\}$ . □

## 1.2. Generalized Inverses.

**Definition 1.9.** Let  $A \in \mathcal{M}_{m,n}$  and  $G \in \mathcal{M}_{n,m}$ . Then  $G$  is said to be a **generalized inverse** of  $A$  if  $AGA = A$ .

**Theorem 1.10.** Let  $A \in \mathcal{M}_{m,n}$ . Suppose that  $\text{rank}(A) = r$ . Then there exists  $P \in \mathcal{M}_{m,m}$ ,  $Q \in \mathcal{M}_{n,n}$ ,  $C \in \mathcal{M}_{r,r}$  such that  $P, Q, C$  are non-singular,  $\text{rank}(C) = r$  and

$$A = P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q$$

**Exercise 1.11.** Let

$$A = P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q$$

as in the previous theorem and  $D \in \mathcal{M}_{r,m-r}$ ,  $E \in \mathcal{M}_{n-r,r}$ ,  $F \in \mathcal{M}_{n-r,m-r}$ . Put

$$G = Q^{-1} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} P^{-1}$$

Then  $G$  is a generalized inverse of  $A$ .

*Proof.*

$$\begin{aligned} AGA &= \left[ P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \right] \left[ Q^{-1} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} P^{-1} \right] \left[ P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \right] \\ &= P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C^{-1} & D \\ E & F \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= P \begin{pmatrix} I & CD \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q \\ &= A \end{aligned}$$

□

**Note 1.12.** The previous exercise and theorem guarantee the existence of a generalized inverse for all matrices. We will take  $G = A^-$  to mean that  $G$  is a generalized inverse of  $A$ . Unless otherwise specified,  $A^-$  will refer to a generic generalized inverse of  $A$ , that is, unless otherwise specified, any statement about  $A^-$  will apply to all generalized inverses of  $A$ .

**Theorem 1.13.** Let  $A \in \mathcal{M}_{m,n}$ . Suppose that  $\text{rank}(A) = r$ . Let  $P \in \mathcal{M}_{m,m}$ ,  $Q \in \mathcal{M}_{n,n}$  permutation matrices and  $C \in \mathcal{M}_{r,r}$ . Suppose that  $\text{rank}(C) = r$  and  $PAQ = \begin{pmatrix} C & D \\ E & F \end{pmatrix}$ . Then

$$Q \begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix} P = A^-.$$

**Exercise 1.14.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $(X^T)^- = (X^-)^T$ .

*Proof.*

$$\begin{aligned} X^T(X^-)^T X^T &= (X X^- X)^T \\ &= X^T \end{aligned}$$

□

**Exercise 1.15.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $\mathcal{C}(XX^-) = \mathcal{C}(X)$ .

*Proof.* Clearly  $\mathcal{C}(XX^-) \subset \mathcal{C}(X)$ . Let  $b \in \mathcal{C}(X)$ . Then there exists  $a \in \mathbb{R}^n$  such that  $Xa = b$ . Then

$$\begin{aligned} XX^-b &= XX^-Xa \\ &= Xa \\ &= b \end{aligned}$$

So  $b \in \mathcal{C}(XX^-)$ . Thus  $\mathcal{C}(X) \subset \mathcal{C}(XX^-)$  and  $\mathcal{C}(X) = \mathcal{C}(XX^-)$  □

**Exercise 1.16.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $\mathcal{N}(X) = \mathcal{N}(X^-X)$

*Proof.* From the previous exercise, we have that

$$\begin{aligned} \mathcal{N}(X) &= \mathcal{C}(X^T)^\perp \\ &= \mathcal{C}(X^T(X^T)^-)^\perp \\ &= \mathcal{C}(X^T(X^-)^T)^\perp \\ &= \mathcal{C}((X^-X)^T)^\perp \\ &= \mathcal{N}(X^-X) \end{aligned}$$

□

**Exercise 1.17.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $X^- = (X^T X)^- X^T$ .

*Proof.* By definition,  $X^T X (X^T X)^- X^T X = X^T X$ . A previous exercise implies that  $X(X^T X)^- X^T X = X$ . Thus  $X^- = (X^T X)^- X^T$ . □

### 1.3. Projections.

**Definition 1.18.** Let  $A \in \mathcal{M}_{m,m}$ . Then  $A$  is said to be **idempotent** if  $A^2 = A$ .

**Exercise 1.19.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $XX^-$  and  $X^-X$  are idempotent

*Proof.*

$$\begin{aligned} (XX^-)(XX^-) &= (XX^-X)X^- \\ &= XX^- \end{aligned}$$

The case is similar for  $X^-X$ . □

**Exercise 1.20.** Let  $A \in \mathcal{M}_{m,m}$ . If  $A$  is idempotent, then  $I - A$  is idempotent.

*Proof.* Suppose that  $A$  is idempotent. Then

$$\begin{aligned} (I - A)(I - A) &= I^2 - IA - AI + A^2 \\ &= I - 2A + A \\ &= I - A \end{aligned}$$

□

**Theorem 1.21.** Let  $A \in \mathcal{M}_{m,m}$ . If  $A$  is idempotent, then  $\text{rank}(A) = \text{tr}(A)$ .

**Definition 1.22.** Let  $P \in \mathcal{M}_{m,m}$  and  $S \subset \mathbb{R}^m$  a subspace. Then  $P$  is said to be a **projection matrix** onto  $S$  if

- (1)  $P$  is idempotent
- (2)  $\mathcal{C}(P) \subset S$
- (3) for each  $x \in S$ ,  $Px = x$

**Note 1.23.** In the previous definition, (2) and (3) imply that  $\mathcal{C}(X) = S$ , so to say that  $X$  projects “onto”  $S$  is accurate.

**Exercise 1.24.** Let  $S \subset \mathbb{R}^m$  and  $P, Q$  projection matrices onto  $S$ . Then  $PQ = Q$ .

*Proof.* Let  $x \in \mathbb{R}^m$ . Then  $Qx \in \mathcal{C}(Q) = S$ . So  $PQx = Qx$ . Thus  $PQ = Q$ . □

**Exercise 1.25.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $XX^-$  is a projection onto  $\mathcal{C}(X)$ .

*Proof.* A previous exercises tells us that  $XX^-$  is idempotent. Another previous exercise tells us that  $\mathcal{C}(XX^-) = \mathcal{C}(X)$ . Let  $b \in \mathcal{C}(X)$ . Then there exists  $a \in \mathbb{R}^n$  such that  $Xa = b$ . So

$$\begin{aligned} XX^-b &= XX^-Xa \\ &= Xa \\ &= b \end{aligned}$$

□

**Exercise 1.26.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $I - X^-X$  is a projection onto  $\mathcal{N}(X)$

*Proof.* Since  $X^-X$  is idempotent, so is  $I - X^-X$ . Let  $b \in \mathcal{C}(I - X^-X)$ . Then there exists  $a \in \mathbb{R}^n$  such that  $(I - X^-X)a = b$ . Then

$$\begin{aligned} Xb &= X(I - X^-X)a \\ &= (X - XX^-X)a \\ &= (X - X)a \\ &= 0a \\ &= 0 \end{aligned}$$

So  $\mathcal{C}(I - X^-X) \subset \mathcal{N}(X)$ . Let  $a \in \mathcal{N}(X)$ . Then  $Xa = 0$  and

$$\begin{aligned} (I - X^-X)a &= a - X^-Xa \\ &= a \end{aligned}$$

So for each  $a \in \mathcal{N}(X)$ ,  $(I - X^-X)a = a$ . □

**Exercise 1.27.** Let  $S \subset \mathbb{R}^m$  be a subspace and  $P \in \mathcal{M}_{m,m}$  be a symmetric projection matrix onto  $S$ . Then  $P$  is unique.

*Proof.* Let  $Q \in \mathcal{M}_{m,m}$  be a symmetric projection matrix onto  $S$ . Then

$$\begin{aligned} (P - Q)^T(P - Q) &= P^T P - P^T Q - Q^T P + Q^T Q \\ &= P^2 - PQ - QP + Q^2 \\ &= P - Q - P + Q \\ &= 0 \end{aligned}$$

Thus  $P - Q = 0$  and  $P = Q$ . □

**Definition 1.28.** Let  $X \in \mathcal{M}_{m,n}$ . We define  $P_X$  by

$$P_X = X(X^T X)^- X^T$$

**Exercise 1.29.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $P_X$  is well defined, that is,  $P_X$  is independent of the choice of  $(X^T X)^-$ .

*Proof.* Suppose that  $G, H$  are generalized inverses of  $X^T X$ . By definition, we have

$$\begin{aligned} X^T X G X^T X &= X^T X H X^T X \Rightarrow X G X^T X = X H X^T X \\ &\Rightarrow X^T X G^T X^T = X^T X H X^T \\ &\Rightarrow X G^T X^T = X H X^T \\ &\Rightarrow X G X^T = X H X^T = P_X \end{aligned}$$

□

**Note 1.30.** Recall that  $X^- = (X^T X)^- X^T$ . So that  $P_X = X X^-$  is indeed a projection onto  $\mathcal{C}(X)$ . Recall that  $[(X^T X)^-]^T$  is a generalized inverse of  $(X^T X)^T = (X^T X)$ . Hence  $P_X^T = X[(X^T X)^-]^T X^T = P_X$ . Since  $P_X$  is symmetric, it is the unique symmetric projection onto  $\mathcal{C}(X)$ .

**Exercise 1.31.** Let  $X \in \mathcal{M}_{m,n}$ . Then  $(X^T)^- = X(X^T X)^-$ .

*Proof.* We know that  $P_X X = X$ . Transposing both sides, we get that

$$\begin{aligned} X^T &= X^T P_X \\ &= X^T X (X^T X)^- X^T \end{aligned}$$

So

$$(X^T)^- = X(X^T X)^-$$

□

**Note 1.32.** Recall that  $(X^T)^- = X(X^T X)^-$ . So that  $P_X = (X^T)^- X^T$ . A previous exercises tells us that  $I - P_X$  is a projection on  $\mathcal{N}(X^T)$ . Since  $I - P_X$  is symmetric, it is the unique symmetric projection onto  $\mathcal{N}(X^T)$ .

**Exercise 1.33.** Let  $X_1, X_2 \in \mathcal{M}_{m,n}$ . Suppose that  $\mathcal{C}(X_1) = \mathcal{C}(X_2)^\perp$ . Then  $P_{X_1} P_{X_2} = P_{X_2} P_{X_1} = 0$ .

*Proof.* Since  $I - P_{X_1}$  is the unique symmetric projection onto  $\mathcal{N}(X_1^T) = \mathcal{C}(X_1)^\perp = \mathcal{C}(X_2)$ , we have that  $I - P_{X_1} = P_{X_2}$ . Thus  $P_{X_1} P_{X_2} = P_{X_1} (I - P_{X_1}) = 0$ . Similarly,  $P_{X_2} P_{X_1} = 0$ . □

**Exercise 1.34.** Let  $X \in \mathcal{M}_{m,n}$ . For each  $z \in \mathcal{N}(X^T)$ ,  $P_X z = 0$ .

*Proof.* Let  $z \in \mathcal{N}(X^T)$ . Then  $P_X z = X(X^T X)^- (X^T z) = 0$ . □

**Exercise 1.35.** Let  $X_1, X_2 \in \mathcal{M}_{m,n}$ . If  $\mathcal{C}(X_1) \subset \mathcal{C}(X_2)$ , then  $P_{X_2} - P_{X_1}$  is the unique projection onto  $\mathcal{C}((I - P_{X_1})X_2)$ .

*Proof.* Clearly  $P_{X_2} - P_{X_1}$  is symmetric. Since  $\mathcal{C}(X_1) \subset \mathcal{C}(X_2)$ , we have that  $P_{X_2} P_{X_1} = P_{X_1}$ . Also, by symmetry,

$$\begin{aligned} (P_{X_1} P_{X_2})^T &= P_{X_2}^T P_{X_1}^T \\ &= P_{X_2} P_{X_1} \\ &= P_{X_1} \end{aligned}$$

So  $P_{X_1}P_{X_2} = P_{X_1}^T = P_{X_1}$ . Now we have that

(1)

$$\begin{aligned} (P_{X_2} - P_{X_1})^2 &= (P_{X_2} - P_{X_1})(P_{X_2} - P_{X_1}) \\ &= P_{X_2}^2 + P_{X_1}^2 - P_{X_2}P_{X_1} - P_{X_1}P_{X_2} \\ &= P_{X_2} + P_{X_1} - P_{X_1} - P_{X_1} \\ &= P_{X_2} - P_{X_1} \end{aligned}$$

So  $P_{X_2} - P_{X_1}$  is idempotent.

(2) Let  $x \in \mathbb{R}^m$ . Then there exist unique  $y \in \mathcal{C}(X_2)$  and  $z \in \mathcal{C}(X_2)^\perp = \mathcal{N}(X_2^T)$  such that  $x = y + z$ . So there exists  $e \in \mathbb{R}^n$  such that  $y = X_2e$ . Since  $z \in \mathcal{N}(X_2^T)$ ,  $P_{X_2}z = 0$ . Then

$$\begin{aligned} (P_{X_2} - P_{X_1})x &= P_{X_2}x - P_{X_1}x \\ &= P_{X_2}x - P_{X_1}P_{X_2}x \\ &= y - P_{X_1}y \\ &= X_2e - P_{X_1}X_2e \\ &= (I - P_{X_1})X_2e \\ &\in \mathcal{C}((I - P_{X_1})X_2) \end{aligned}$$

(3) Let  $x \in \mathcal{C}((I - P_{X_1})X_2)$ . Then there exists  $e \in \mathbb{R}^n$  such that  $x = (I - P_{X_1})X_2e$ . So

$$\begin{aligned} (P_{X_2} - P_{X_1})x &= P_{X_2}(I - P_{X_1})x \\ &= P_{X_2}(I - P_{X_1})(I - P_{X_1})X_2e \\ &= P_{X_2}(I - P_{X_1})X_2e \\ &= (P_{X_2} - P_{X_1})X_2e \\ &= P_{X_2}X_2e - P_{X_1}X_2e \\ &= X_2e - P_{X_1}X_2e \\ &= (I - P_{X_1})X_2e \\ &= x \end{aligned}$$

□

#### 1.4. Solving Linear Equations.

**Definition 1.36.** Let  $A \in \mathcal{M}_{m,n}$  and  $b \in \mathbb{R}^m$ . Then the system  $Ax = b$  is said to be **consistent** if  $b \in \mathcal{C}(A)$ .

**Exercise 1.37.** Let  $A \in \mathcal{M}_{m,n}$  and  $G \in \mathcal{M}_{n,m}$ . Then  $G = A^-$  iff for each  $b \in \mathcal{C}(A)$ ,  $Gb$  solves  $Ax = b$ .

*Proof.* Suppose that  $G = A^-$ . Let  $b \in \mathcal{C}(A)$ . Then there exists  $x^* \in \mathbb{R}^n$  such that  $Ax^* = b$ . So

$$\begin{aligned} A(Gb) &= AG(Ax^*) \\ &= (AGA)x^* \\ &= Ax^* \\ &= b \end{aligned}$$

So  $Gb$  solves  $Ax = b$ . Conversely, Suppose that for each  $b \in \mathcal{C}(A)$ ,  $Gb$  solves  $Ax = b$ . Let  $z \in \mathbb{R}^n$ . So  $Az \in \mathcal{C}(A)$ . Then

$$\begin{aligned}(AGA)z &= A[G(Az)] \\ &= Az\end{aligned}$$

Since for each  $z \in \mathbb{R}^n$   $AGAz = Az$ ,  $AGA = A$  and  $G = A^-$ . □

**Exercise 1.38.** Let  $b \in \mathcal{C}(A)$ . Then

$$\{x \in \mathbb{R}^n : Ax = b\} = \{A^-b + (I - A^-A)z : z \in \mathbb{R}^n\}$$

.

*Proof.* Let  $x \in \{A^-b + (I - A^-A)z : z \in \mathbb{R}^n\}$ . Then there exists  $z \in \mathbb{R}^n$  such that  $x = A^-b + (I - A^-A)z$ . Since  $(I - A^-A)$  is a projection onto  $\mathcal{N}(A)$ ,

$$\begin{aligned}Ax &= AA^-b \\ &= b\end{aligned}$$

So  $x \in \{x \in \mathbb{R}^n : Ax = b\}$ . Conversely, let  $x \in \{x \in \mathbb{R}^n : Ax = b\}$ . Then

$$\begin{aligned}x &= A^-(Ax) + (x - A^-Ax) \\ &= A^-(b) + (I - A^-A)x \\ &\in \{A^-b + (I - A^-A)z : z \in \mathbb{R}^n\}\end{aligned}$$

□

### 1.5. Moore-Penrose Pseudoinverse.

**Theorem 1.39. Singular Value Decomposition:**

Let  $A \in \mathcal{M}_{m,n}$ . Suppose that  $\text{rank}(A) = r$ . Then there exist  $U \in \mathcal{M}_{m,m}$ ,  $V \in \mathcal{M}_{n,n}$ , and  $D_0 \in \mathcal{M}_{r,r}$  such that

- (1)  $A = U \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} V^T$
- (2)  $U^T U = I$
- (3)  $V^T V = I$
- (4)  $D_0 = \text{diagonal}(d_1, d_2, \dots, d_r)$  with  $d_1 \geq d_2 \geq \dots \geq d_r > 0$

**Note 1.40.** Put  $D = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{m,n}$

- (1) Since  $D_0$  is symmetric,  $D^T = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$
- (2) Since  $D_0$  is diagonal,  $D_0^{-1}$  is also diagonal and symmetric

**Definition 1.41.** Let  $A \in \mathcal{M}_{m,m}$  and  $A^+ \in \mathcal{M}_{n,m}$ . Then  $A^+$  is said to be a **Moore-Penrose pseudoinverse** of  $A$  if

- (1)  $AA^+A = A$
- (2)  $A^+AA^+ = A^+$
- (3)  $AA^+$  is symmetric
- (4)  $A^+A$  is symmetric



**Note 1.42.** We have that  $P_X = XX^+ = X(X^T X)^- X^T$ .

**Exercise 1.43.** Let  $A \in \mathcal{M}_{m,n}$  and  $S, T \in \mathcal{M}_{n,m}$ . If  $S$  and  $T$  are  $m$ - $p$  pseudoinverses of  $A$ , then  $S = T$ .

*Proof.* Suppose that  $S, T$  satisfy properties (1)-(4). Then

$$\begin{aligned}
 S &= SAS \\
 &= (SA)^T S \\
 &= A^T S^T S \\
 &= (ATA)^T S^T S \\
 &= A^T T^T A^T S^T S \\
 &= (TA)^T (SA)^T S \\
 &= (TA)(SA)S \\
 &= TA(SAS) \\
 &= TAS
 \end{aligned}$$

and

$$\begin{aligned}
 T &= TAT \\
 &= T(AT)^T \\
 &= TT^T A^T \\
 &= TT^T (ASA)^T \\
 &= TT^T A^T S^T A^T \\
 &= T(AT)^T (AS)^T \\
 &= T(AT)(AS) \\
 &= (TAT)AS \\
 &= TSA
 \end{aligned}$$

So  $S = T$

□

**Exercise 1.44.** Let  $A \in \mathcal{M}_{m,n}$  have singular value decomposition  $A = UDV^T = U \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} V^T$ .

Define  $D^+ = \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$ . Then  $D^+$  is the  $m$ - $p$  pseudoinverse of  $D$ .

*Proof.*

(1)

$$\begin{aligned}
DD^+D &= \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \\
&= D
\end{aligned}$$

(2) Similar to (1).

(3)

$$\begin{aligned}
(DD^+)^T &= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}^T \\
&= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \\
&= DD^+
\end{aligned}$$

(4) Similar to (3).

□

**Exercise 1.45.** Let  $A \in \mathcal{M}_{m,n}$  have singular value decomposition  $A = UDV^T$ . So  $A^T \in \mathcal{M}_{n,m}$  has singular value decomposition  $A^T = VD^TU^T$ . Then  $(D^T)^+ = (D^+)^T$

*Proof.* Since  $D^T = \begin{pmatrix} D_0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}$ , we have that  $(D^T)^+ = \begin{pmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{pmatrix} = (D^+)^T$  □

**Exercise 1.46.** Let  $A \in \mathcal{M}_{m,n}$  have singular value decomposition  $A = UDV^T$ . Define  $A^+ = VD^+U^T$ . Then  $A^+$  is the  $m$ - $p$  pseudoinverse of  $A$ .

*Proof.* (1)

$$\begin{aligned}
AA^+A &= (UDV^T)(VD^+U^T)(UDV^T) \\
&= UDD^+DV^T \\
&= UDV^T \\
&= A
\end{aligned}$$

(2) Similar to (1)

(3)

$$\begin{aligned}
(AA^+)^T &= [(UDV^T)(VD^+U^T)]^T \\
&= (UDD^+U^T)^T \\
&= U(DD^+)^TU^T \\
&= UDD^+U^T \\
&= (UDV^T)(VD^+U^T) \\
&= AA^+
\end{aligned}$$

(4) Similar to (3). □

**Exercise 1.47.** Let  $A \in \mathcal{M}_{m,n}$  have singular value decomposition  $A = UDV^T$ . Then  $(A^T)^+ = (A^+)^T$ .

*Proof.*

$$\begin{aligned} (A^T)^+ &= [(UDV^T)^T]^+ \\ &= (VD^T U^T)^+ \\ &= U(D^T)^+ V^T \\ &= U(D^+)^T V^T \\ &= (VD^+ U^T)^T \\ &= (A^+)^T \end{aligned}$$

□

**Exercise 1.48.** Let  $A \in \mathcal{M}_{m,n}$ . Then there exists a unique matrix  $A^+ \in \mathcal{M}_{n,m}$  such that  $A^+$  is the  $m$ - $p$  pseudoinverse of  $A$ .

*Proof.* The existence of and uniqueness of  $A^+$  are shown in the previous exercises. □

**Exercise 1.49.** Let  $A \in \mathcal{M}_{m,m}$ . Then  $(A^+)^+ = A$ .

*Proof.* We observe that  $A$  satisfies properties (1)–(4) for  $A^+$ . By uniqueness,  $(A^+)^+ = A$ . □

**Exercise 1.50.** Let  $A \in \mathcal{M}_{m,n}$  and  $b \in \mathcal{C}(A)$ . Put  $S = \{x \in \mathbb{R}^n : Ax = b\}$ . Then

$$\|A^+b\| = \min_{x \in S} \|x\|$$

.

*Proof.* Let  $x \in S$ . A previous exercise tells us that there exists  $z \in \mathbb{R}^n$  such that  $x = A^+b + (I - A^+A)z$ . Then

$$\begin{aligned} \|x\|^2 &= \|A^+b + (I - A^+A)z\|^2 \\ &= (A^+b + (I - A^+A)z)^T (A^+b + (I - A^+A)z) \\ &= \|A^+b\|^2 - 2z^T (I - A^+A)^T (A^+b) + \|(I - A^+A)z\|^2 \\ &= \|A^+b\|^2 - 2z^T (I - A^+A)A^+b + \|(I - A^+A)z\|^2 \\ &= \|A^+b\|^2 + \|(I - A^+A)z\|^2 \\ &\geq \|A^+b\|^2 \end{aligned}$$

□

## 1.6. Differentiation.

**Definition 1.51.** Let  $Q : \mathbb{R}^n \rightarrow \mathbb{R}$  given by  $b \mapsto Q(b)$ . Suppose that  $Q \in C^1(\mathbb{R}^n)$ . We define

$$\frac{\partial Q}{\partial b} = \begin{pmatrix} \frac{\partial Q}{\partial b_1} \\ \vdots \\ \frac{\partial Q}{\partial b_n} \end{pmatrix}$$

**Exercise 1.52.** Let  $a, b \in \mathbb{R}_n$  and  $A \in \mathcal{M}_{n,n}$ . Then

$$(1) \quad \frac{\partial a^T b}{\partial b} = a$$

$$(2) \quad \frac{\partial b^T A b}{\partial b} = (A + A^T)b$$

*Proof.*

(1) Since

$$a^T b = \sum_{i=1}^n a_i b_i$$

We have that

$$\frac{\partial a^T b}{\partial b_i} = a_i$$

and therefore

$$\frac{\partial a^T b}{\partial b} = a$$

(2) Since

$$\begin{aligned} b^T A b &= \sum_{i=1}^n b_i \sum_{j=1}^n A_{i,j} b_j \\ &= \sum_{i=1}^n \sum_{j=1}^n b_i A_{i,j} b_j \end{aligned}$$

The terms containing  $b_i$  are

$$A_{i,i} b_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^n (A_{i,j} + A_{j,i}) b_i b_j$$

This implies that

$$\begin{aligned} \frac{\partial b^T A b}{\partial b_i} &= 2A_{i,i} b_i + \sum_{\substack{j=1 \\ j \neq i}}^n (A_{i,j} + A_{j,i}) b_j \\ &= \sum_{j=1}^n (A_{i,j} + A_{j,i}^T) b_j \\ &= [(A + A^T)b]_i \end{aligned}$$

So

$$\frac{\partial b^T A b}{\partial b} = (A + A^T)b$$

□

## 1.7. Quadratic Forms and Eigendecomposition.

**Definition 1.53.** Let  $A \in \mathcal{M}_{n,n}$ . Then  $A$  is said to be **positive semi-definite** if for each  $x \in \mathbb{R}^n$ ,

$$x^T A x \geq 0$$

**Definition 1.54.** Let  $A \in \mathcal{M}_{n,n}$ . Then  $A$  is said to be **positive-definite** if for each  $x \in \mathbb{R}^n$ ,  $x \neq 0$  implies that

$$x^T A x > 0$$

**Exercise 1.55.** Let  $A \in \mathcal{M}_{n,n}$ . If  $A$  is positive-definite, then  $A$  is invertible.

**Exercise 1.56.** Let  $A \in \mathcal{M}_{n,n}$ . Then  $A$  is invertible iff for each eigenvalue  $\lambda$  of  $A$ ,  $\lambda \neq 0$ .

*Proof.* Suppose that  $A$  is invertible. Let  $x \in \mathbb{R}^n$ . Suppose that  $Ax = 0$ . Then  $x = 0$ . So  $x$  is not an eigenvector of  $A$ . So 0 is not an eigenvalue of  $A$ . Conversely. Suppose that  $A$  is not invertible. Then there exists  $x \in \ker A$  such that  $x \neq 0$ . Then  $Ax = 0 = 0x$ . So 0 is an eigenvalue of  $A$ .  $\square$

*Proof.* Suppose that  $A$  is positive definite. Let  $x \in \ker A$ . Suppose that  $x \neq 0$ . Then  $x^T A x = x^T 0 = 0$ , which is a contradiction. Hence  $\ker A = \{0\}$ . So  $\text{rank}(A) = n$  and  $A$  is invertible.  $\square$

**Exercise 1.57.** Let  $A \in \mathcal{M}_{n,n}$ . If  $A$  is positive semi-definite (respectively positive definite), then the eigenvalues of  $A$  are nonnegative (respectively positive).

*Proof.* Let  $\lambda \in \mathbb{C}$  be an eigenvalue of  $A$  and  $x \in \mathbb{R}^n$  a corresponding eigenvector. Then  $x \neq 0$ . So  $x^T x \geq 0$ . If  $A$  is positive semi-definite, then

$$\begin{aligned} 0 &\leq x^T A x \\ &= \lambda x^T x \end{aligned}$$

Hence  $\lambda \geq 0$ . The case is similar for  $A$  positive definite.  $\square$

**Definition 1.58.** Let  $U \in \mathcal{M}_{n,n}$ . Then  $U$  is said to be **orthogonal** if  $U^T U = U U^T = I$

**Theorem 1.59.** Let  $A \in \mathcal{M}_{n,n}$  be a symmetric matrix and  $\lambda_1, \dots, \lambda_n$  the eigenvalues of  $A$ . Then

- (1)  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$
- (2) for  $i, j \in \{1, \dots, n\}$  if  $i \neq j$  and  $x_i, x_j$  are eigenvectors corresponding to  $\lambda_i, \lambda_j$  respectively, then  $x_i^T x_j = 0$ .
- (3) there exist  $U, D \in \mathcal{M}_{n,n}$  such that  $U$  is orthogonal,  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  and  $A = U D U^T$ .

**Note 1.60.** We will be dealing with covariance matrices which are positive semi-definite symmetric matrices and thus have nonnegative eigenvalues.

## 2. THE LINEAR MODEL

### 2.1. Model Description.

**Definition 2.1.** Given  $y \in \mathbb{R}_m$  a vector of observed responses to the matrix  $X \in \mathcal{M}_{m,n}$  of observed inputs, we will consider the model

$$y = Xb + e$$

where  $b \in \mathbb{R}_n$  is a vector of unknown parameters and  $e \in \mathbb{R}^m$  is a random vector of unobserved errors with zero mean.

**Definition 2.2.** For a parameter vector  $b \in \mathbb{R}^n$ , we have that  $e = y - Xb$ . For this reason,  $e$  is called the **residual vector** or simply the “residuals”.

**Note 2.3.** The goal will be to find a parameter vector  $b \in \mathbb{R}^n$  that makes the residuals to be as small as possible.

## 2.2. Least Squares Optimization.

**Definition 2.4.** We define the **cost function**,  $Q : \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$\begin{aligned} Q(b) &= \|y - Xb\|^2 \\ &= (y - Xb)^T(y - Xb) \end{aligned}$$

**Definition 2.5.** Let  $b \in \mathbb{R}^n$ . Then  $b$  is said to be a **least squares solution** for the model if

$$Q(b) = \inf_{c \in \mathbb{R}^n} Q(c)$$

**Exercise 2.6.** If  $b$  is a least squares solution for the model, then  $X^T Xb = X^T y$ .

*Proof.* Suppose that  $b$  is a least squares solution for the model, then  $Q$  has a global minimum at  $b$ . Since  $Q$  is convex in  $b$ , this global minimum is also a local minimum. Thus

$$\frac{\partial Q}{\partial b}(b) = 0$$

By definition,

$$\begin{aligned} Q(b) &= y^T y - y^T Xb - b^T X^T y + b^T X^T Xb \\ &= y^T y - 2y^T Xb + b^T X^T Xb \end{aligned}$$

Thus

$$\begin{aligned} 0 &= \frac{\partial Q}{\partial b}(b) \\ &= -2X^T y + 2X^T Xb \end{aligned}$$

Hence  $X^T Xb = X^T y$ . □

**Definition 2.7.** For  $y \in \mathbb{R}^m$  and  $X \in \mathcal{M}_{m,n}$ , we define the **normal equation** to be

$$X^T Xb = X^T y$$

**Exercise 2.8.** The normal equation is consistent.

*Proof.* We have that  $X^T y \in \mathcal{C}(X^T) = \mathcal{C}(X^T X)$ . □

**Exercise 2.9.** Let  $b \in \mathbb{R}^n$ . Then  $b$  is a least squares solution for the model iff  $b$  satisfies the normal equation.

*Proof.* The previous exercises tells us that if  $b$  is a least squares solution for the model, then  $b$  satisfies the normal equation. Conversely, suppose that  $b$  satisfies the normal equation.

Then

$$\begin{aligned}
 Q(c) &= (y - Xc)^T(y - Xc) \\
 &= (y - Xb + Xb - Xc)^T(y - Xb + Xb - Xc) \\
 &= (y - Xb)^T(y - Xb) - (y - Xb)^T(X(b - c)) - (b - c)^T X^T(y - Xb) + (b - c)^T X^T(X(b - c)) \\
 &= Q(b) - 2(b - c)^T X^T(y - Xb) + \|X(b - c)\|^2 \\
 &= Q(b) + \|X(b - c)\|^2
 \end{aligned}$$

Thus  $b$  minimizes  $Q$ . □

**Exercise 2.10.** Let  $b \in \mathbb{R}^n$  be a least squares solution for the model. Then  $\|y\|^2 = \|Xb\|^2 + \|e\|^2$

*Proof.* Since  $b$  satisfies the normal equation, we have that  $X^T(y - Xb) = 0$ . Thus

$$\begin{aligned}
 Xb \cdot e &= b^T X^T e \\
 &= b^T X^T(y - Xb) \\
 &= b^T 0 \\
 &= 0
 \end{aligned}$$

So  $Xb$  and  $e$  are orthogonal. Therefore

$$\begin{aligned}
 \|y\|^2 &= \|Xb + e\|^2 \\
 &= \|Xb\|^2 + \|e\|^2
 \end{aligned}$$

□

### 2.3. Estimation.

**Note 2.11.** In what follows we are considering the model  $y = Xb + e$  with  $y, e \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^p$ ,  $X \in \mathcal{M}_{n,p}$  and  $\mathbb{E}[e] = 0$ .

**Definition 2.12.** Let Then  $\lambda \in \mathbb{R}^p$ . The function  $t(y)$  is said to be a linear unbiased estimator for the function  $f(b) = \lambda^T b$  if there exists  $a \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  such that  $t(y) = c + a^T y$  and for each  $b \in \mathbb{R}^p$ ,  $\mathbb{E}[t(y)] = \lambda^T b$ .

**Exercise 2.13.** Let Then  $\lambda \in \mathbb{R}^p$  and  $a \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$ . Suppose that  $t(y) = c + a^T y$  is an unbiased linear estimator for  $f(b) = \lambda^T b$ . Then  $c = 0$  and  $\lambda = X^T a$ .

*Proof.* We have that for each  $b \in \mathbb{R}^p$ ,

$$\begin{aligned}
 \lambda^T b &= \mathbb{E}[c + a^T y] \\
 &= c + a^T \mathbb{E}[y] \\
 &= c + a^T Xb
 \end{aligned}$$

Taking  $b = 0$ , we get that  $c = 0$ . So for each  $b \in \mathbb{R}^p$ ,  $\lambda^T b = a^T Xb$ . This implies that  $\lambda^T = a^T X$  and  $\lambda = X^T a$ . □

**Definition 2.14.** Let  $\lambda \in \mathbb{R}^p$ . Then the function  $f(b) = \lambda^T b$  is said to be **linearly estimable** if there exists a linear, unbiased estimator for  $f(b)$ . Equivalently,  $f(b) = \lambda^T b$  is linearly estimable if there exists  $a \in \mathbb{R}^n$  such that for each  $b \in \mathbb{R}^p$   $\mathbb{E}[a^T y] = \lambda^T b$

**Exercise 2.15.** Let  $\lambda \in \mathbb{R}^p$ . Then the following are equivalent:

- (1)  $f(b) = \lambda^T b$  is linearly estimable
- (2)  $\lambda \in \mathcal{C}(X^T)$
- (3) for each  $G \in X^-$  of  $X$ ,  $\lambda^T = \lambda^T G X$
- (4) there exists  $G \in X^-$  of  $X$  such that  $\lambda^T = \lambda^T G X$

$f(b) = \lambda^T b$  is linearly estimable iff  $\lambda \in \mathcal{C}(X^T)$ .

*Proof.* (1)  $\Rightarrow$  (2)

Suppose that  $f(b)$  is linearly estimable. Then there exists  $a \in \mathbb{R}^n$  such that for each  $b \in \mathbb{R}^p$   $\mathbb{E}[a^T y] = \lambda^T b$ . Then for each  $b \in \mathbb{R}^p$ ,

$$\lambda^T b = a^T \mathbb{E}[y] = a^T X b$$

Hence  $\lambda^T = a^T X$  and  $X^T a = \lambda$ . So  $\lambda \in \mathcal{C}(X^T)$ .

(2)  $\Rightarrow$  (3)

Suppose that  $\lambda \in \mathcal{C}(X^T)$ . Let  $G \in X^-$ . Then  $G^T \in (X^T)^-$ . Since  $\lambda \in \mathcal{C}(X^T)$ , there exists  $a \in \mathbb{R}^n$  such that  $X^T a = \lambda$ . A previous exercise tells us that there exists  $z \in \mathbb{R}^n$  such that

$$a = G^T \lambda + (I - G^T X^T) z$$

So

$$\begin{aligned} \lambda &= X^T a \\ &= X^T [G^T \lambda + (I - G^T X^T) z] \\ &= X^T G^T \lambda \end{aligned}$$

Hence  $\lambda^T = \lambda^T G X$ .

(3)  $\Rightarrow$  (4)

Trivial.

(4)  $\Rightarrow$  (1)

Suppose that there exists  $G \in X^-$  such that  $\lambda^T = \lambda^T G X$ . Choose  $a = G^T \lambda \in \mathbb{R}^n$ . Let  $b \in \mathbb{R}^p$ . Then

$$\begin{aligned} \mathbb{E}[a^T y] &= a^T \mathbb{E}[y] \\ &= \lambda^T G \mathbb{E}[y] \\ &= \lambda^T G X b \\ &= \lambda^T b \end{aligned}$$

So  $f(b) = \lambda^T b$  is linearly estimable. □

**Definition 2.16.** Let  $\hat{b} \in \mathbb{R}^p$  be a least squares solution and  $\lambda \in \mathbb{R}^n$ . Then  $\hat{f} = \lambda^T \hat{b}$  is said to be a least squares estimator of  $f(b) = \lambda^T b$ .

**Exercise 2.17.** Let  $\hat{b} \in \mathbb{R}^p$  be a least squares solution and  $\lambda \in \mathbb{R}^n$ . Then  $\hat{f} = \lambda^T \hat{b}$  is the unique least squares estimator of  $f(b) = \lambda^T b$  iff  $f(b)$  is linearly estimable.



*Proof.* Suppose that  $f(b) = \lambda^T b$  is linearly estimable. Then  $\lambda \in \mathcal{C}(X^T)$ . So there exists  $a \in \mathbb{R}^n$  such that  $\lambda^T = a^T X$ . Let  $b'$  be a least squares solution. Then there exists  $z \in \mathbb{R}^p$  such that

$$b' = (X^T X)^- X^T y + (I - (X^T X)^- (X^T X))z$$

Then

$$\begin{aligned} \lambda^T b' &= \lambda^T \left[ (X^T X)^- X^T y + (I - (X^T X)^- (X^T X))z \right] \\ &= a^T X (X^T X)^- X^T y + a^T X (I - (X^T X)^- X^T X)z \\ &= a^T P_X y + a^T (X - P_X X)z \\ &= a^T P_X y \end{aligned}$$

In particular,  $\lambda^T b' = a^T P_X y = \lambda^T \hat{b}$ .

Conversely, suppose that  $\hat{f} = \lambda^T \hat{b}$  is the unique least squares estimator of  $f(b) = \lambda^T b$ . Then for each  $z \in \mathbb{R}^p$ ,

$$\lambda^T \hat{b} = \lambda^T (X^T X)^- X^T y + \lambda^T (I - (X^T X)^- X^T X)z$$

So for each  $z \in \mathbb{R}^p$ ,

$$\lambda^T (X^T X)^- X^T y + \lambda^T (I - (X^T X)^- X^T X)z = 0$$

and thus

$$\lambda^T (I - (X^T X)^- X^T X) = 0$$

Therefore

$$\lambda^T = \lambda^T (X^T X)^- X^T X$$

Transposing both sides, we obtain that

$$\lambda = X^T X [(X^T X)^-]^T \lambda \in \mathcal{C}(X^T)$$

So  $f(b) = \lambda^T b$  is linearly estimable □

**Exercise 2.18.** Let  $\lambda \in \mathbb{R}^p$  and  $\hat{b} \in \mathbb{R}^p$  a least squares solution. If  $f(b) = \lambda^T b$  is linearly estimable, then the unique least squares estimator  $\hat{f} = \lambda^T \hat{b}$  of  $f(b)$  is a linear unbiased estimator of  $f(b)$ .

*Proof.* Suppose that  $f(b) = \lambda^T b$  is linearly estimable. Then there exists  $a \in \mathbb{R}^n$  such that  $\lambda^T = a^T X$ . The previous exercise tells us that

$$\lambda^T \hat{b} = \lambda^T (X^T X)^- X^T y$$

Thus for each  $b \in \mathbb{R}^p$ ,

$$\begin{aligned} \mathbb{E}[\lambda^T \hat{b}] &= \mathbb{E}[\lambda^T (X^T X)^- X^T y] \\ &= \lambda^T (X^T X)^- X^T \mathbb{E}[y] \\ &= \lambda^T (X^T X)^- X^T X b \\ &= a^T X (X^T X)^- X^T X b \\ &= a^T P_X X b \\ &= a^T X b \\ &= \lambda^T b \end{aligned}$$

□

## 2.4. Imposing Restrictions for a Unique Solution.

**Definition 2.19.** Let  $X \in \mathcal{M}_{n,p}$  with  $\text{rank}(X) = r$ , let  $s = p - r$  and  $C \in \mathcal{M}_{s,p}$  with  $\text{rank}(C) = s$  and  $\mathcal{C}(X^T) \cap \mathcal{C}(C^T) = \{0\}$  and let  $y \in \mathbb{R}^n$ . We consider the system

$$\begin{pmatrix} X^T X \\ C \end{pmatrix} b = \begin{pmatrix} X^T y \\ 0 \end{pmatrix}$$

or equivalently the system

$$\begin{pmatrix} X \\ C \end{pmatrix} b = \begin{pmatrix} P_X y \\ 0 \end{pmatrix}$$

These systems are the **restricted normal equations with restrictions  $C$** .

**Note 2.20.** Requiring  $\text{rank}(C) = s$  means that the rows of  $C$  (i.e. the restrictions) are linearly independent. To have a unique solution to

$$\begin{pmatrix} X \\ C \end{pmatrix} b = \begin{pmatrix} P_X y \\ 0 \end{pmatrix}$$

we must have

$$\mathcal{N}\left(\begin{pmatrix} X \\ C \end{pmatrix}\right) = \{0\}$$

or equivalently,

$$\mathcal{C}\left(\begin{pmatrix} X^T & C^T \end{pmatrix}\right) = \mathbb{R}^p$$

Since  $\text{rank}(X^T) = \text{rank}(X) = r$ , we have that  $\mathcal{C}\left(\begin{pmatrix} X^T & C^T \end{pmatrix}\right) = \mathbb{R}^p$  iff  $\mathcal{C}(X^T) \cap \mathcal{C}(C^T) = \{0\}$ .

**Exercise 2.21.** Under the assumptions for the restricted normal equations, the following systems are equivalent:

(1)

$$\begin{pmatrix} X^T X \\ C \end{pmatrix} b = \begin{pmatrix} X^T y \\ 0 \end{pmatrix}$$

(2)

$$\begin{pmatrix} X^T X \\ C^T C \end{pmatrix} b = \begin{pmatrix} X^T y \\ 0 \end{pmatrix}$$

(3)

$$(X^T X + C^T C)b = X^T y$$

*Proof.* (1)  $\Rightarrow$  (2) We need to show that for each  $b \in \mathbb{R}^p$   $Cb = 0$  implies that  $C^T Cb = 0$ . This is immediate since  $\mathcal{N}(C^T C) = \mathcal{N}(C)$ .

(2)  $\Rightarrow$  (3) Let  $b \in \mathbb{R}^p$  be a solution to system (1). Then we have that

$$(X^T X + C^T C)b = X^T Xb + C^T Cb = X^T y + 0 = X^T y$$

(3)  $\Rightarrow$  (1) Suppose that  $(X^T X + C^T C)b = X^T y$ . This implies that  $C^T Cb = X^T(y - Xb)$ . So

$$C^T Cb \in \mathcal{C}(C^T C) \cap \mathcal{C}(X^T) = \mathcal{C}(C^T) \cap \mathcal{C}(X^T) = \{0\}$$

Hence  $b \in \mathcal{N}(C^T C) = \mathcal{N}(C)$ . So  $Cb = 0$  and  $X^T Xb = (X^T X + C^T C)b = X^T y$ , orquivalently,

$$\begin{pmatrix} X^T X \\ C \end{pmatrix} b = \begin{pmatrix} X^T y \\ 0 \end{pmatrix}$$

□

**Exercise 2.22.** Under the assumptions for the restricted normal equations, we have the following:

- (1)  $X^T X + C^T C$  is invertible
- (2)  $(X^T X + C^T C)^{-1} X^T y$  is the unique solution to  $X^T Xb = X^T y$  and  $Cb = 0$ .
- (3)  $(X^T X + C^T C)^{-1}$  is a generalized inverse of  $X^T X$
- (4)  $C(X^T X + C^T C)^{-1} X^T = 0$
- (5)  $C(X^T X + C^T C)^{-1} C^T = I$

*Proof.*

(1)

$$\begin{aligned} \mathbb{R}^p &= \mathcal{C} \left( \begin{pmatrix} X^T & C^T \end{pmatrix} \right) \\ &= \mathcal{C} \left( \begin{pmatrix} X^T & C^T \end{pmatrix} \begin{pmatrix} X \\ C \end{pmatrix} \right) \\ &= \mathcal{C}(X^T X + C^T C) \end{aligned}$$

Since  $X^T X + C^T C \in \mathcal{M}_{p,p}$  and  $\text{rank}(X^T X + C^T C) = p$ , we have that  $X^T X + C^T C$  is invertible.

- (2) Put  $b = (X^T X + C^T C)^{-1} X^T y$ . Then  $(X^T X + C^T C)b = X^T y$ . A previous exercise tells us that  $b$  is a solution to the system

$$\begin{pmatrix} X^T X \\ C \end{pmatrix} b = \begin{pmatrix} X^T y \\ 0 \end{pmatrix}$$

which implies that  $X^T Xb = X^T y$  and  $Cb = 0$ .

- (3) From (2), we know that

$$X^T X [(X^T X + C^T C)^{-1} X^T y] = X^T y$$

Since  $y \in \mathbb{R}^n$  is arbitrary, we have

$$X^T X (X^T X + C^T C)^{-1} X^T = X^T$$

Multiplying both sides on the right by  $X$  tells us that  $(X^T X + C^T C)^{-1}$  is a generalized inverse of  $X^T X$ .

- (4) From (2), we know that

$$C(X^T X + C^T C)^{-1} X^T y = 0$$

Since  $y \in \mathbb{R}^n$  is arbitrary,

$$C(X^T X + C^T C)^{-1} X^T = 0$$

(5)

□

## 2.5. Constrained Parameter Space.

**Definition 2.23.** Let  $P \in \mathcal{M}_{p,q}$  and  $\delta \in \mathbb{R}^q$ . Suppose that  $P$  has full column rank. We define the **constrained parameter space**  $\mathcal{T} = \{b \in \mathbb{R}^p : P^T b = \delta\}$ .

**Note 2.24.** Since  $P$  has full column rank,  $\mathcal{C}(P^T) = \mathbb{R}^q$  and for each  $\delta \in \mathbb{R}^q$ ,  $P^T b = \delta$  is consistent. We now fix  $P, \delta$  so that  $\mathcal{T}$  is fixed.

**Definition 2.25.** Let  $\lambda \in \mathbb{R}^p$ . The function  $t(y)$  is said to be a **linear unbiased estimator in  $\mathcal{T}$**  for  $f(b) = \lambda^T b$  if there exists  $a \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  such that  $t(y) = c + a^T y$  and for each  $b \in \mathcal{T}$ ,  $\mathbb{E}[t(y)] = \lambda^T b$ .

**Definition 2.26.** Let  $\lambda \in \mathbb{R}^p$ . The function  $f(b) = \lambda^T b$  is said to be **linearly estimable in  $\mathcal{T}$**  if there exists a linear unbiased estimator in  $\mathcal{T}$  for  $f(b) = \lambda^T b$ . Equivalently  $\lambda^T b$  is linearly estimable in  $\mathcal{T}$  if there exist  $a \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  such that for each  $b \in \mathcal{T}$ ,  $\mathbb{E}[c + a^T y] = \lambda^T b$ .

**Theorem 2.27.** Let  $\lambda \in \mathbb{R}^p$  and  $a \in \mathbb{R}^n$ . Then  $t(y) = c + a^T y$  is a linear unbiased estimator for  $f(b) = \lambda^T b$  iff if there exists  $d \in \mathbb{R}^q$  such that  $\lambda = X^T a + P d$  and  $c = d^T \delta$ .

**Definition 2.28.** We define the **normal equations with restrictions  $\mathcal{T}$**  to be

$$\begin{pmatrix} X^T X & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} b \\ \theta \end{pmatrix} = \begin{pmatrix} X^T y \\ \delta \end{pmatrix}$$

**Theorem 2.29.** We have the following:

- (1) The restricted normal equations are consistent.
- (2) Let  $\hat{b}$  be the first component of a solution to the restricted normal equations. Then  $Q(\hat{b}) = \min_{b \in \mathcal{T}} Q(b)$ .
- (3) Let  $\hat{b}$  be the first component of a solution to the restricted normal equations and  $b \in \mathcal{T}$ . Then  $Q(b) = Q(\hat{b})$  iff  $b$  is the first component of a solution of to the restricted normal equations.

## 2.6. The Gauss-Markov Model.

**Definition 2.30.** Let  $X \in \mathcal{M}_{n,p}$ ,  $y \in \mathbb{R}^n$ . We consider the model  $y = Xb + e$  where  $\mathbb{E}[e] = 0$ ,  $\text{Var}(e) = \sigma^2 I_n$ . This model is called the **Gauss-Markov model**. Note that  $\mathbb{E}[y] = Xb$  and  $\text{Var}(y) = \sigma^2 I$ .

**Theorem 2.31.** Let  $a, c \in \mathbb{R}^n$ ,  $A \in \mathcal{M}_{p,n}$  and  $y$  a random vector in  $\mathbb{R}^n$ . Then

- (1)  $\mathbb{E}[a^T y] = a^T \mathbb{E}[y]$
- (2)  $\text{Var}(a^T y) = a^T \text{Var}(y) a$
- (3)  $\text{Cov}(a^T y, c^T y) = a^T \text{Var}(y) c$
- (4)  $\text{Var}(Ay) = A^T \text{Var}(y) A$

**Exercise 2.32.** Let  $\lambda^T \in \mathbb{R}^p$  and  $\hat{b} \in \mathbb{R}^p$  a least squares solution. Suppose that  $f(b) = \lambda^T b$  is linearly estimable. Then the unique least squares estimator  $\hat{f} = \lambda^T \hat{b}$  satisfies

$$\text{Var}(\hat{f}) = \sigma^2 \lambda^T (X^T X)^{-} \lambda$$

*Proof.* Uniqueness of  $\hat{f}$  tells us that  $\hat{f} = \lambda^T (X^T X)^{-} X^T y$ . A previous exercise tells us that for each gen. inv.  $X^-$  of  $X$ ,  $\lambda^T = \lambda^T X^- X$ . Recall that  $(X^T X)^{-} X^T$  is a gen. inv. of  $X$ .

Then

$$\begin{aligned}
 \text{Var}(\hat{f}) &= \text{Var}(\lambda^T (X^T X)^- X^T y) \\
 &= \lambda^T (X^T X)^- X^T \text{Var}(y) (\lambda^T (X^T X)^- X^T)^T \\
 &= \sigma^2 \lambda^T (X^T X)^- X^T (\lambda^T (X^T X)^- X^T)^T \\
 &= \sigma^2 \lambda^T (X^T X)^- \left( \lambda^T (X^T X)^- X^T X \right)^T \\
 &= \sigma^2 \lambda^T (X^T X)^- (\lambda^T)^T \\
 &= \sigma^2 \lambda^T (X^T X)^- \lambda
 \end{aligned}$$

□

**Exercise 2.33.** Let  $\lambda \in \mathbb{R}^p$ . Suppose that  $f(b) = \lambda^T b$  is linearly estimable. Then  $\hat{f} = \lambda^T \hat{b}$  is the minimum variance linear unbiased estimator for  $f(b)$ .

*Proof.* Let  $t(y) = c + a^T y$  be a linear unbiased estimator for  $f(b) = \lambda^T b$ . Recall that  $c = 0$  and  $\lambda = X^T a$ ,  $\hat{f} = \lambda^T (X^T X)^- X^T y$  and for each generalized inverse  $X^-$  of  $X$ ,  $\lambda^T X^- X = \lambda^T$ . Then

$$\begin{aligned}
 \text{Var}(t(y)) &= \text{Var}(a^T y) \\
 &= \text{Var}(\hat{f} + (a^T y - \hat{f})) \\
 &= \text{Var}(\hat{f}) + \text{Var}(a^T y - \hat{f}) + 2\text{Cov}(\hat{f}, a^T y - \hat{f})
 \end{aligned}$$

Now

$$\begin{aligned}
 \text{Cov}(\hat{f}, a^T y - \hat{f}) &= \text{Cov}(\lambda^T (X^T X)^- X^T y, a^T y - \lambda^T (X^T X)^- X^T y) \\
 &= \lambda^T (X^T X)^- X^T \text{Var}(y) \left[ a^T - \lambda^T (X^T X)^- X^T \right]^T \\
 &= \sigma^2 \lambda^T (X^T X)^- X^T \left[ a^T - \lambda^T (X^T X)^- X^T \right]^T \\
 &= \sigma^2 \lambda^T (X^T X)^- \left[ a^T X - \lambda^T (X^T X)^- X^T X \right]^T \\
 &= \sigma^2 \lambda^T (X^T X)^- \left[ a^T X - \lambda^T (X^T X)^- X^T X \right]^T \\
 &= \sigma^2 \lambda^T (X^T X)^- \left[ a^T X - \lambda^T \right]^T \\
 &= \sigma^2 \lambda^T (X^T X)^- (X^T a - \lambda) \\
 &= 0
 \end{aligned}$$

Hence  $\text{Var}(t(y)) = \text{Var}(\hat{f}) + \text{Var}(a^T y - \hat{f}) \geq \text{Var}(\hat{f})$

□

**Theorem 2.34.**

- (1) For each  $A, B \in \mathcal{M}_{n,n}$  and  $\alpha \in \mathbb{R}$ ,  $\text{tr}(\alpha A + B) = \alpha \text{tr}(A) + \text{tr}(B)$ .
- (2) For each  $A \in \mathcal{M}_{n,p}$  and  $B \in \mathcal{M}_{p,n}$ ,  $\text{tr}(AB) = \text{tr}(BA)$ .
- (3) For each random matrix  $Z \in \mathcal{M}_{n,n}$ ,  $\mathbb{E}[\text{tr}(Z)] = \text{tr}(\mathbb{E}[Z])$ .

**Exercise 2.35.** Let  $z \in \mathbb{R}^p$  be a random vector. Suppose that  $\mathbb{E}[z] = \mu$  and  $\text{Var}(z) = \Sigma$ . Then for each  $A \in \mathcal{M}_{p,p}$ ,

$$\mathbb{E}[z^T A z] = \mu^T A \mu + \text{tr}(A \Sigma)$$

*Proof.* Note that

$$\mathbb{E}[z^T A z] = \mathbb{E}[(z - \mu)^T A (z - \mu)] + \mathbb{E}[\mu^T A (z - \mu)] + \mathbb{E}[z^T A \mu]$$

Observe that

$$\begin{aligned} \mathbb{E}[(z - \mu)^T A (z - \mu)] &= \mathbb{E}[\text{tr}((z - \mu)^T A (z - \mu))] \\ &= \mathbb{E}[\text{tr}((A(z - \mu)(z - \mu)^T))] \\ &= \text{tr}(\mathbb{E}[(A(z - \mu)(z - \mu)^T)]) \\ &= \text{tr}(A \mathbb{E}[(z - \mu)(z - \mu)^T]) \\ &= \text{tr}(A \Sigma) \end{aligned}$$

and that

$$\begin{aligned} \mathbb{E}[\mu^T A (z - \mu)] &= \mathbb{E}[\mu^T A (z - \mu)] \\ &= \mu^T A \mathbb{E}[z - \mu] \\ &= 0 \end{aligned}$$

and that

$$\begin{aligned} \mathbb{E}[z^T A \mu] &= \mathbb{E}[z^T] A \mu \\ &= \mu^T A \mu \end{aligned}$$

Thus  $\mathbb{E}[z^T A z] = \mu^T A \mu + \text{tr}(A \Sigma)$ . □

**Definition 2.36.** Put  $\hat{e} = y - \hat{y} = (I - P_X)y$ . Then the **sum of squares error**,  $SSE$ , is defined to be  $SSE = \hat{e}^T \hat{e} = y^T (I - P_X)y$ .

**Exercise 2.37.** Let  $r = \text{rank}(X)$ . Define

$$\hat{\sigma}^2 = \frac{SSE}{n - r}$$

Then  $\hat{\sigma}^2$  is an unbiased estimator for  $\sigma^2$ .

*Proof.* The previous exercise tells us that

$$\begin{aligned} \mathbb{E}[SSE] &= \mathbb{E}[y^T (I - P_X)y] \\ &= b^T X^T (I - P_X) X b + \sigma^2 \text{tr}(I - P_X) \\ &= \sigma^2 \text{tr}(I - P_X) \\ &= \sigma^2 \text{rank}(I - P_X) \\ &= \sigma^2 \text{nullity}(X^T) \\ &= \sigma^2 (n - r) \end{aligned}$$

So  $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ . □

## 2.7. The Aitken Model.

**Definition 2.38.** Let  $X \in \mathcal{M}_{n,p}$ ,  $y \in \mathbb{R}^n$  and  $V \in \mathcal{M}_{n,n}$ . We consider the model  $y = Xb + e$  where  $\mathbb{E}[e] = 0$ ,  $\text{Var}(e) = \sigma^2 V$ . This model is called the **Aitken model**. Note that  $\mathbb{E}[y] = Xb$  and  $\text{Var}(y) = \sigma^2 V$ .

**Definition 2.39.** Let  $R \in \mathcal{M}_{n,n}$ . Suppose that  $R$  is invertible and  $RV R^T = I$  or equivalently,  $V = (R^T R)^{-1}$ . We define the **transformed Aitken model** by  $z = Ry$ ,  $U = RX$ ,  $f = Re$  so that

$$z = Ub + f$$

Note that

$$\mathbb{E}[z] = RXb = Ub$$

and

$$\text{Var}(f) = R\text{Var}(e)R^T = \sigma^2 RV R^T = \sigma^2 I$$

**Definition 2.40.** Under the transformed Aitken model, we can look for solutions  $b \in \mathbb{R}^p$  to the normal equations

$$U^T U b = U^T z$$

When we transform back to the Aitken model, we have the **Aitken equations**

$$X^T V^{-1} X b = X^T V^{-1} y$$

We denote a solution to the Aitken equations by  $\hat{b}_{GLS}$  and a solution to the normal equations by  $\hat{b}_{OLS}$

## 3. DISTRIBUTION THEORY

### 3.1. Introduction.

**Definition 3.1.** Let  $x \in \mathbb{R}^p$  be a random vector. Define  $m_x : \mathbb{R}^p \rightarrow [0, \infty]$  by

$$m_x(t) = \mathbb{E}[e^{t^T x}]$$

We call  $m_x$  the **moment generating function of  $X$** .

**Theorem 3.2.** Let  $x_1, x_2 \in \mathbb{R}^p$  be random vectors. Then  $F_{x_1} = F_{x_2}$  iff  $m_{x_1} = m_{x_2}$ .

**Theorem 3.3.** Let  $x_1 \in \mathbb{R}^{p_1}, \dots, x_n \in \mathbb{R}^{p_n}$  be random vectors. Put  $p = \sum_{i=1}^n p_i$ . For  $t \in \mathbb{R}^p$ , we can partition  $t$  as  $t = (t_1^T, \dots, t_n^T)^T$  where  $t_1 \in \mathbb{R}^{p_1}, \dots, t_n \in \mathbb{R}^{p_n}$ . Put  $x = (x_1^T, \dots, x_n^T)^T$ . Then  $x_1, \dots, x_n$  are independent iff for each  $t \in \mathbb{R}^p$ ,  $m_x(t) = \prod_{i=1}^n m_{x_i}(t_i)$ .

### 3.2. Multivariate Normal.

**Definition 3.4.** Let  $x \in \mathbb{R}^p$  be a random vector,  $\mu \in \mathbb{R}^p$  and  $V \in \mathcal{M}_{p,p}$  be symmetric and positive semi-definite. Then  $x$  is said to have a **multivariate normal distribution with mean  $\mu$  and covariance matrix  $V$** , denoted  $x \sim N_p(\mu, V)$ , if for each  $t \in \mathbb{R}^p$ ,  $m_x(t) = e^{t^T \mu + \frac{1}{2} t^T V t}$ .

**Exercise 3.5.** Let  $x \sim N_p(\mu, V)$ ,  $a \in \mathbb{R}^q$  and  $B \in \mathcal{M}_{q,p}$ . Define the random vector  $y \in \mathbb{R}^q$  by  $y = a + Bx$ . Then  $y \sim N_q(a + B\mu, BV B^T)$ .

*Proof.* for  $t \in \mathbb{R}^q$ , we have that

$$\begin{aligned}
m_y(t) &= E[e^{t^T y}] \\
&= E[e^{t^T(a+Bx)}] \\
&= E[e^{t^T a} + t^T Bx] \\
&= e^{t^T a} E[e^{t^T Bx}] \\
&= e^{t^T a} m_x(B^T t) \\
&= e^{t^T a} e^{t^T B\mu + \frac{1}{2} t^T BVB^T t} \\
&= e^{t^T a + t^T B\mu + \frac{1}{2} t^T BVB^T t} \\
&= e^{t^T(a+B\mu) + \frac{1}{2} t^T BVB^T t}
\end{aligned}$$

So  $y \sim N_q(a + B\mu, BVB^T)$ . □

**Exercise 3.6.** Let  $x \in \mathbb{R}^p$  be a multivariate normal random vector. Then any subvector of  $x$  is a multivariate normal random vector.

*Proof.* Let  $x = (x_1, \dots, x_p)^T$ . Suppose that  $x \sim N_p(\mu, V)$ . Let  $x' = (x_{i_1}, \dots, x_{i_k})$  be a subvector of  $x$ . So  $i_1 < \dots < i_k$  and  $i_1, \dots, i_k \in \{1, \dots, p\}$ . Choose a matrix  $B \in \mathcal{M}_{q,p}$  such that  $x' = Bx$ . Then  $x' \sim N_q(B\mu, BVB^T)$ . □

**Theorem 3.7.** Let  $x = (x_1^T, \dots, x_n^T)^T$ ,  $\mu = (\mu_1^T \dots, \mu_n^T)^T \in \mathbb{R}^p$  and  $V = \begin{pmatrix} V_{1,1} & \dots & V_{1,n} \\ & \ddots & \\ V_{n,1} & \dots & V_{n,n} \end{pmatrix} \in$

$\mathcal{M}_{p,p}$  where  $x_i, \mu_i \in \mathbb{R}^{p_i}$ ,  $V_{i,j} \in \mathcal{M}_{p_i, p_j}$  and  $\sum_{i=1}^n p_i = p$ . If  $x \sim N_p(\mu, V)$  then  $x_1, \dots, x_n$  are independent iff for each  $i, j \in \{1, \dots, n\}$ ,  $i \neq j$  implies that  $V_{i,j} = 0$ .

**Exercise 3.8.** Let  $z = (z_1, \dots, z_p) \in \mathbb{R}^p$  be a random vector. Then  $z \sim N_p(0, I)$  iff for each  $i = 1, \dots, p$ ,  $z_i \sim N_1(0, 1)$  and  $z_1, \dots, z_n$  are independent.

*Proof.* Suppose that  $z \sim N_p(0, I)$ . Since  $z_i = e_i^T z$ , the previous results tells us that for each  $i = 1, \dots, p$ ,  $z_i \sim N_1(0, 1)$  and  $z_1, \dots, z_n$  are independent. Conversely, suppose that for each  $i = 1, \dots, p$ ,  $z_i \sim N_1(0, 1)$  and  $z_1, \dots, z_n$  are independent. Then for each  $t \in \mathbb{R}^p$ ,

$$\begin{aligned}
m_z(t) &= \prod_{i=1}^p m_{z_i}(t_i) \\
&= \prod_{i=1}^p e^{\frac{1}{2} t_i^2} \\
&= e^{\frac{1}{2} t^T t}
\end{aligned}$$

Thus  $z \sim N_p(0, I)$ . □

**Exercise 3.9.** Let  $x \sim N_p(\mu, V)$ ,  $a_1, a_2 \in \mathbb{R}^q$ ,  $B_1, B_2 \in \mathcal{M}_{q,p}$ ,  $y_1 = a_1 + B_1 x$  and  $y_2 = a_2 + B_2 x$ . Then  $y_1, y_2$  are independent iff  $B_1 V B_2^T = 0$ .



*Proof.* Put  $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ ,  $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$  and  $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$ . Since  $y = a + Bx$  We know that  $y \sim N_{2q}(a + B\mu, BV B^T)$ . Observe that  $BV B^T = \begin{pmatrix} B_1 V B_1^T & B_1 V B_2^T \\ B_2 V B_1^T & B_2 V B_2^T \end{pmatrix}$  A previous result tells us that  $y_1, y_2$  are independent iff  $B_1 V B_2^T = 0$   $\square$

### 3.3. Chi-Square.

**Definition 3.10.** Let  $u$  be a random variable,  $p \in \mathbb{N}$  and  $\phi \geq 0$ . Then  $u$  is said to have a  $\chi^2$  **distribution with  $p$  degrees of freedom and noncentrality parameter  $\phi$** , denoted  $u \sim \chi_p^2(\phi)$ , if for each  $t \in \mathbb{R}$ ,

$$m_u(t) = (1 - 2t)^{p/2} e^{2\phi t/(1-2t)}$$

If  $\phi = 0$ , we say that  $u$  has a (central) chi-square distribution with  $p$  degrees of freedom, denoted  $u \sim \chi_p^2$ .

**Exercise 3.11.** Let  $z \sim N_p(0, I)$ . Then  $z^T z \sim \chi_p^2$ .

*Proof.* One can show that  $m_{z_i^2}(t) = (1 - 2t)^{1/2}$ . By independence

$$m_{z^T z}(t) = m_{\sum_{i=1}^p z_i^2}(t) = (1 - 2t)^{p/2}$$

$\square$

**Exercise 3.12.** Let  $u_1 \sim \chi_{p_1}^2(\phi_1), \dots, u_n \sim \chi_{p_n}^2(\phi_n)$ . Put  $u = \sum_{i=1}^n u_i$ ,  $p = \sum_{i=1}^n p_i$  and  $\phi = \sum_{i=1}^n \phi_i$ . If  $u_1, \dots, u_n$  are independent, then  $u \sim \chi_p^2(\phi)$ .

*Proof.* By independence, we have that for each  $t \in \mathbb{R}$ ,

$$\begin{aligned} m_u(t) &= \prod_{i=1}^n m_{u_i}(t) \\ &= \prod_{i=1}^n (1 - 2t)^{p_i/2} e^{2\phi_i t/(1-2t)} \\ &= (1 - 2t)^{p/2} e^{2\phi t/(1-2t)} \end{aligned}$$

$\square$

**Theorem 3.13.** Let  $x \in \mathbb{R}^p$  be a random vector,  $\mu \in \mathbb{R}^p$  and  $V \in \mathcal{M}_{p,p}$  positive definite. If  $x \sim N_p(\mu, V)$ , then

$$x^T V^{-1} x \sim \chi_p^2(\mu^T V^{-1} \mu/2)$$

**Definition 3.14.** Let  $p_1, p_2 \in \mathbb{N}$ ,  $\phi \geq 0$ ,  $u_1 \sim \chi_{p_1}^2(\phi)$  and  $u_2 \sim \chi_{p_2}^2$ . Then  $\frac{u_1/p_1}{u_2/p_2}$  is said to have a  $F$  **distribution with noncentrality parameter  $\phi$  and  $p_1, p_2$  degrees of freedom**, denoted  $\frac{u_1/p_1}{u_2/p_2} \sim F_{p_1, p_2}(\phi)$ .

**Definition 3.15.** Let  $\mu \in \mathbb{R}$ ,  $k \in \mathbb{N}$ ,  $u \sim N_1(\mu, 1)$  and  $v \sim \chi_k^2$ . Then  $u/\sqrt{v/k}$  is said to have a  $t$  **distribution with noncentrality parameter  $\mu$  and  $k$  degrees of freedom**, denoted  $u/\sqrt{v/k} \sim t_k(\mu)$ .

**3.4. Quadratic Forms.**

**Exercise 3.16.** *Let  $\mu \in \mathbb{R}^p$  and  $A, V \in \mathcal{M}_{p,p}$ . Suppose that  $A$  is symmetric,  $V$  is nonsingular,  $AV$  is idempotent and  $\text{rank}(AV) = s$ . Let  $x \sim N_p(\mu, V)$ . Then  $x^T A x \sim \chi_p^2(\mu A \mu^T / 2)$*

**Exercise 3.17.**