Introduction

These are my notes from the book of Harville [**?**]. This are probably very esoteric notes, I noted whatever was important for me. That is, however basic, anything that I didn't know found a place here. It is also a summary of what I gained from this book. My coverage of chapters will probably be insanely unbalanced.

Chapter 1

Matrices

Preliminaries and basics of matrix algebra are presented here. But one has to go through to become familiar with the notation that will follow (*e.g.* $\mathbf{J}_{mn}$ is an $m \times n$ matrix of ones). Plus going through the entire text and solving the exercises helped me to start and become familiar with *proving* linear algebra. Exercise 8 is particularly interesting.

Some notes:

- $\mathbf{J}_{mn}\mathbf{J}_{np} = n\mathbf{J}_{mp}$

- **L1.3.1** If $\mathbf{A}, \mathbf{B}$ are upper (lower) triangular, then (i) $\mathbf{AB}$ is upper (lower) triang. (ii) the ith diagonal of $\mathbf{AB}$ is $a_{ii}b_{ii}$

1.0.1 Redo in future

- Exercise 4.a–c

- Exercise 8

Chapter 2

Submatrices and Partitioned Matrices

A basic topic that I wasn't sufficiently familiar is covered in this chapter (see title).

2.1 Definitions and some properties

- *Principle submatrix* is a submatrix of $\mathbf{A}$ that is obtained by striking out the rows and columns with the same indices.

- *Leading principal submatrix* is an $r \times r$ submatrix of $\mathbf{A}$ obtained by striking out the last $n - r$ columns and rows.

- A *partitioning* of a matrix or *partitioned matrix* is obtained by putting horizontal/vertical lines between its various columns or rows. The only requirement is that the lines shouldn't be *staggered* (see #15).

Notes

- The fact that multiplication of partitioned matrices is similar to that of the regular matrices (see 2.7) is very intuitive. Proves useful when proving thing in future chapters.

Interesting identities/lemmas/theorems

- (2.9): $\mathbf{AB} = \sum_{k=1}^{c} \mathbf{A}_k \mathbf{B}_k$, where $\mathbf{A}, \mathbf{B}$ are appropriately defined partitionings (see #19), i.e. $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_c)$, $B = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \ldots \\ \mathbf{B}_c \end{bmatrix}$. Particularly interesting when $\mathbf{A}_i$ and $\mathbf{B}_j$ are cols and rows of $\mathbf{A}$ and $\mathbf{B}$ respectively, i.e. $\mathbf{C} = \mathbf{AB} = \sum_{i=1}^{c} \mathbf{a}_i \mathbf{b}_i'$.

- **L2.3.2** For any two $m \times n$ matrices $\mathbf{A}, \mathbf{B}$, $\mathbf{A} = \mathbf{B} \iff \mathbf{Ax} = \mathbf{Bx}$ for every $n \times 1$ vector $\mathbf{x}$

Important concepts

- Block-diagonal matrix

- Block-triangular matrix

2.1.1 Redo in future

- Exercise 4

Chapter 3

Linear Dependence and Independence

Very short chapter, but its lemmas are very useful, they can be used extensively in proofs that come later in the chapters. Lemma 3.2.1 and 3.2.2 are fundamental and they should be practically memorized. Lemma 3.2.4 is an elegant one, it seems like it can be the key to some complicated proofs.

3.1 Definitions

- A nonempty finite set of matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_k\}$ is called *linearly dependent* if $\exists$ scalars $x_1, \ldots, x_k$ such that

$$x_1\mathbf{A}_1 + \ldots + x_k\mathbf{A}_k = \mathbf{0}$$

3.2 Notes

- **L3.2.1** A set of $m \times n$ matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_k\}$ is linearly dependent iff there is at least one matrix that is expressible as the linear combination of others.

- **L3.2.2** A set of $m \times n$ matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_k\}$, the first of which is nonnull, is linearly dependent fif at least one of the matrices is expressible as the linear combination of the preceding ones.

- **L3.2.4** Let $\{\mathbf{A}_i\}_{i=1}^k$ and $\{\mathbf{C}_i\}_{i=1}^r$ be sets of matrices and $\mathbf{x}_1, \ldots, \mathbf{x}_r$ vectors such that

$$\mathbf{C}_j = x_{1j}\mathbf{A}_1 + \ldots + x_{kj}\mathbf{A}_k,$$

  where $\mathbf{x}_j = (x_{1j}, \ldots, x_{kj})$. If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ and $\mathbf{A}_1, \ldots, \mathbf{A}_k$ are linearly independent, then $\mathbf{C}_1, \ldots, \mathbf{C}_r$ are linearly independent. If $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are linearly dependent, then $\mathbf{C}_1, \ldots, \mathbf{C}_r$ are linearly dependent.z

3.3 Redo in future

Notes

- Given $\mathbf{A}_1, \ldots, \mathbf{A}_k$. Let $\mathbf{C}_j = x_{1j}\mathbf{A}_1 + \ldots x_{kj}\mathbf{A}_k$ for $j = 1, \ldots, r$ and $\mathbf{x}_j = (x_{1j}, \ldots, x_{kj})$. If $\mathbf{A}_1, \ldots, \mathbf{A}_k$ are linearly independent and $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are linearly independent, then $\mathbf{C}_1, \ldots, \mathbf{C}_r$ are linearly independent. If $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are linearly dependent, then $\mathbf{C}_1, \ldots, \mathbf{C}_r$ are linearly dependent.

- Lemma 3.2.1

- Lemma 3.2.2

- Lemma 3.2.4

- Exercise 2

Chapter 4

Linear Space: Row and Column Spaces

Very useful chapter, it's title doesn't do justice. Fundamental concepts of rank, dimension, basis, linear space, subspace, column/row space are introduced and they are made concrete with many lemmas and theorems.

4.1   Definitions and basic properties

- *Column space* of an $m \times n$ matrix, $\mathcal{C}(\mathbf{A})$, is the set of all m-dimensional col vectors that can be expressed as the linear combination of columns of $\mathbf{A}$, *i.e.* $\mathcal{C}(\mathbf{A}) = \{\sum_{i=1}^{n} x_i \mathbf{a}_i : \forall x_i \in \mathcal{R}\}$.

- *Row space* $\mathcal{R}(\mathbf{A})$ is defined similarly as $\mathcal{R}(\mathbf{A}) = \{\sum_{i=1}^{m} x_i \mathbf{a}'_i : \forall x_i \in \mathcal{R}\}$.

- Simple ways to ascertain a column space.

  - **L4.1.1.** For any $m \times n$ $\mathbf{A}$, $\mathbf{y} \in \mathcal{C}(\mathbf{A}) \iff \mathbf{y}' \in \mathcal{R}(\mathbf{A}')$
  - **L4.2.1.** Given $m \times n$ $\mathbf{A}$. For any subspace $\mathcal{U}$ of $\mathcal{R}^m$, $\mathcal{C}(\mathbf{A}) \in \mathcal{U} \iff$ every column of $\mathbf{A}$ belongs to $\mathcal{V}$. Similarly, for any subspace $\mathcal{V}$ of $\mathcal{R}^n$, $\mathcal{R}(\mathbf{A}) \in \mathcal{V} \iff$ every row of $\mathbf{A}$ belongs to $\mathcal{U}$. Similarly, for any

- *Linear space.* A nonempty set of $m \times n$ is a linear space if (1) $\mathbf{A}, \mathbf{B} \in \mathcal{V} \implies \mathbf{A} + \mathbf{B} \in \mathcal{V}$ and (2) $\mathbf{A} \in \mathcal{V} \implies k\mathbf{A} \in \mathcal{V}$.

- *Subspace.* A subset $\mathcal{U}$ of a linear space $\mathcal{V}$ is a *subspace* if $\mathcal{U}$ itself is a linear space.

- *Span.* The span of a set of matrices $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_k\}$ is the set of all the linear combinations of $\mathcal{A}$, denoted with $\text{span}(\mathcal{A})$. $\mathcal{A}$ is said to span a space $\mathcal{V}$ if $\text{span}(\mathcal{A}) = \mathcal{V}$

- *Basis* for a linear space $\mathcal{V}$ is a set of linearly independent matrices that spans $\mathcal{V}$.

- *Dimension.* The number of matrices in a basis for a lin. space $\mathcal{V}$ is called the dimension.

  - $\dim(\mathcal{R}^{m \times n}) = mn$
  - $\dim(\mathcal{U}) \leq mn$ for any $\mathcal{U} \subset \mathcal{R}^{m \times n}$

- *Column rank and row rank.* The column (or row) rank of a matrix is the dimension of its column (or row) space.

- *Rank* is the column or row rank of an $m \times n$ matrix $\mathbf{A}$; the two latter ranks are the same anyway (T4.4.1).

  - $\text{rank}(\mathbf{A}) \leq m$ and $\text{rank}(\mathbf{A}) \leq n$ (L4.3.3)
  - For any $n \times p$ mat $\mathbf{F}$, $\text{rank}(\mathbf{AF}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{A}F) \leq \text{rank}(\mathbf{F})$ (C4.4.5)
  - $\text{rank}(\mathbf{A}, \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$ and $\text{rank}\begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix} \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$ (L4.5.7)
  - $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}, \mathbf{B})$ (L4.5.8)
  - $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$ (C4.5.9)
  - $\text{rank}\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} = \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$ (L4.5.11)
  - Repeated application of the above leads to $\text{rank}\,[\text{diag}(\mathbf{A}_1, \dots \mathbf{A}_k)] = \text{rank}(\mathbf{A}_1) + \dots \text{rank}(\mathbf{A}_k)$

- *Nonsingular* matrix is an $m \times m$ matrix $\mathbf{A}$ with rank m.

4.2 Important Lemmas/Theorems/Equations

- **Regarding the column/row spaces**.

  - **L4.2.2**. $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A}) \iff \exists \mathbf{F}$ such that $\mathbf{B} = \mathbf{AF}$. Similarly, $\mathcal{R}(\mathbf{C}) \subset \mathcal{R}(\mathbf{A}) \iff \exists \mathbf{L}$ such that $\mathbf{C} = \mathbf{LA}$.
  - **T4.4.6**. Given $\mathbf{A}, \mathbf{B}$. If $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A})$ and $\mathrm{rank}(\mathbf{B}) = \mathrm{rank}(\mathbf{A})$, then $\mathcal{C}(\mathbf{B}) = \mathcal{C}(\mathbf{A})$. Similar for row space.

- **Regarding the ranks and singularity/nonsingularity.**

  - **T4.4.1** The row rank of any matrix equals its column rank.
  - **T4.4.2**. Let $\mathbf{A}$ be an $m \times n$ matrix of row rank r and column rank c. Then, there exists an $m \times c$ mat $\mathbf{B}$ and a $c \times n$ mat $\mathbf{L}$ s.t. $\mathbf{A} = \mathbf{BL}$.
  - **T4.4.4** Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be $m \times n$, $m \times p$ and $q \times n$. Then, $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A}) \implies \mathrm{rank}(\mathbf{B}) \le \mathrm{rank}(\mathbf{C})$ and $\mathcal{R}(\mathbf{C}) \subset \mathcal{R}(\mathbf{A}) \implies \mathrm{rank}(\mathbf{C}) \le \mathrm{rank}(\mathbf{A})$.
  - Let $\mathbf{A}$ be $m \times n$ of rank r. Then, $\exists \, m \times r \, \mathbf{B}$ and $r \times n \, \mathbf{T}$ s.t. $\mathbf{A} = \mathbf{BT}$. Moreover, $\mathrm{rank}(\mathbf{B}) = \mathrm{rank}(\mathbf{T}) = r$.
  - **T4.4.10**. Any matrix of rank r contains an $r \times r$ nonsingular submatrix.
  - **L4.4.11**. Any symmetric matrix of rank r contains an $r \times r$ nonsingular principal submatrix.

- **Regarding partitioned matrices and sums of matrices.** For any $\mathbf{A}, \mathbf{B}, \mathbf{C}$ with respective sizes $m \times n, m \times p, q \times n$,

  - **C4.5.2**: $\mathrm{rank}(\mathbf{A}) \le \mathrm{rank}(\mathbf{A}, \mathbf{B})$; $\mathrm{rank}(\mathbf{B}) \le \mathrm{rank}(\mathbf{A}, \mathbf{B})$; $\mathrm{rank}(\mathbf{A}) \le \mathrm{rank} \begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix}$ and $\mathrm{rank}(\mathbf{C}) \le \mathrm{rank} \begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix}$
  - **L4.5.4**: $\mathcal{C}(\mathbf{A}, \mathbf{B}) = \mathcal{C}(\mathbf{A}, \mathbf{B} - \mathbf{AL})$. Similar result applies to row spaces.

- Lemma 4.2.2 is easily the most important result in the chapter. Many results in the chapter can be proved directly through this lemma.

- Corollary 4.2.3 also very useful.

- Theorem 4.3.2

- Corollary 4.3.3

- Theorem 4.3.4

- Lemma 4.3.5

- Theorem 4.3.7

- Theorem 4.4.2

- The restatements of column/row space results in terms of rank results in Section 4.4.b

- Theorem 4.4.8

- Lemma 4.5.1

- Lemma 4.5.4

- Lemma 4.5.7

- Corollary 4.5.9

- Theorem 4.4.10

4.3 Redo

- All listed above

- Theorem 4.4.4

- Theorem 4.4.9

- Theorem 4.4.10

- Lemma 4.5.11

- Theorem 4.4.1

- Exercise 4

- Exercise 5

- Exercise 6

- Exercise 7

- Exercise 9

- Exercise 10

- Exercise 11a

4.4   Important concepts

- Linear Space

- Row/Column space

- Basis

- Linear Space

- Rank

- Spanning set

- Dimension

Chapter 5

Trace of a (Square) Matrix

The chapter gives some basic identities about traces, and then a very important equivalence for finding whether a matrix is null, $\mathbf{0}$.

Notes

- Basic trace identities and some properties (Section 5.2 and 5.3)
  - $\text{tr}(k\mathbf{A}) = k\text{tr}(\mathbf{A})$
  - $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
  - $\text{tr}(\mathbf{A}'\mathbf{A}) \geq 0$
  - $\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$
  - $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{B}'\mathbf{A}') = \text{tr}(\mathbf{BA}) = \text{tr}(\mathbf{A}'\mathbf{B}')$
  - The last leads to allowing any cyclical rotations: $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$
    (It is easy to remember that only cyclicals are allowed—the matrices have to be multipliable)
  - **L5.3.1**: $\mathbf{A} = \mathbf{0} \iff \text{tr}(\mathbf{A}'\mathbf{A}) = 0$
    C5.3.2: $\mathbf{A} = \mathbf{0} \iff \mathbf{A}'\mathbf{A} = \mathbf{0}$
    C5.3.3: $\mathbf{AB} = \mathbf{AC} \iff \mathbf{A}'\mathbf{AB} = \mathbf{A}'\mathbf{AC}$ and (transposed) $\mathbf{BA}' = \mathbf{CA}' \iff \mathbf{BA}'\mathbf{A} = \mathbf{A}'\mathbf{AC}$

Redo

- Prove Corollary 5.3.2
- Prove Corollary 5.3.3
- Exercise 3

Chapter 6

Geometrical Considerations

Geometry is very important because it allows for an intuitive understanding of many linear algebra concepts. There are two opinions about those who teach and write about linear statistical models. Those who advocate the "geometrical approach" emphasize its elegance and intuitiveness, and those who advocate a more "algebraic approach" highlight its rigour and the fact that it's more suggestive of computational approaches.

Important concepts

- Inner (or dot) Product (the "usual" one): $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}'\mathbf{y}$ for vecs $\mathbf{A}' \cdot \mathbf{B} = \text{tr}(\mathbf{AB})$ for mats

- Norm: $||\mathbf{x}|| = (\mathbf{x}' \cdot \mathbf{x})^{1/2}$

- Angle: $\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||}$ (similar for matrices)

- Distance: $||\mathbf{A} - \mathbf{B}||$

- Orthogonality

Notes

- Schwarz Inequality: $|\mathbf{A} \cdot \mathbf{B}| \leq ||\mathbf{A}|| \, ||\mathbf{B}||$

- Gram-Schmidt Orthogonalization: $x_{ij} = \frac{\mathbf{A}_j \cdot \mathbf{B}_i}{\mathbf{B}_i \cdot \mathbf{B}_i}$

- Any linear space has an orthonormal basis (T6.4.3): Follows from T4.3.9 and the fact that we can apply Gram-Schmidt Orthogonalization to any set of bases

- QR decomposition for (full-column matrix): $\mathbf{A} = \mathbf{BX} = (\mathbf{BD})(\mathbf{EX}) = \mathbf{QR}$, where cols of $\mathbf{Q}$ are orthonormal and $\mathbf{R}$ is upper triangular; cols of $\mathbf{B}$ are only orthogonal and $\mathbf{D}, \mathbf{E}$ are diagonal matrices such that (clearly) $\mathbf{D}^{-1} = \mathbf{E}$ (See Exercise 5 for generalization of QR decomposition to column rank deficient matrices.)

- The four necessary properties for qualifying as inner product:

    1. $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$
    2. $\mathbf{A} \cdot \mathbf{A} \geq 0$ with equality holding iff $\mathbf{A} = 0$
    3. $(k\mathbf{A}) \cdot \mathbf{B} = k(\mathbf{A} \cdot \mathbf{B})$
    4. $(\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} = (\mathbf{A} \cdot \mathbf{C}) + (\mathbf{B} \cdot \mathbf{C})$

Redo

- Schwarz Inequality

- Exercise 1

- Exercise 5

Chapter 7

Consistency and Compatibility

This chapter is about the two important concepts in the title that relate to whether a linear system has solution(s). Finding the solution(s) is the topic of Chapter 11.

Important Concepts

- Linear system (in $\mathbf{x}$) is: $\mathbf{Ax} = \mathbf{B}$. $\mathbf{A}$ is the *coefficient matrix*, B is the *right side* and any x that satisfies the equality is a solution.

- A linear system is called *homogeneous* if $\mathbf{B} = \mathbf{0}$ and *non-homogeneous* otherwise.

- A system is called *consistent* if it has one or more solutions.

Notes

- **T7.2.1** gives four necessary and sufficient conditions for consistency. All four can be derived from the first condition: $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A})$ (and then using L4.2.2, L4.5.1 and C4.5.2).

- **T7.2.2**: A system is *consistent* if $\mathbf{A}$ has full row rank.

- A linear system is *compatible* if $\mathbf{k}'\mathbf{B} = \mathbf{0}$ for every $\mathbf{k}'$ such that $\mathbf{k}'\mathbf{A} = \mathbf{0}$

- **T7.3.1** is the the chapter's main theorem that connects the two main concepts: A linear system is compatible iff it is consistent

- **T7.4.1** connects the chapter with the Trace chapter; specifically with C5.3.3. This leads to important equivalences between $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}$ in terms of column/row spaces and ranks (see Section 7.4.b).

- **L7.4.5**. For any $\mathbf{A}$, $\mathcal{C}(\mathbf{A}'\mathbf{A}) = \mathcal{C}(\mathbf{A}')$, $\mathcal{R}(\mathbf{A}'\mathbf{A}) = \mathcal{R}(\mathbf{A})$ and $\mathrm{rank}(\mathbf{A}'\mathbf{A}) = \mathrm{rank}(\mathbf{A})$.

Redo

- Prove T7.2.1 and T7.2.2

- Prove T7.3.1

- Prove T7.4.1

- Prove all T's/C's in Section 7.4.b: this will help with consolidating parts of Chapter 4.

- Prove T7.4.8

- Exercise 1b

Chapter 8

Inverse Matrices

The chapter introduces the (left, right) inverses and their basic properties. Orthogonal matrices are also introduced. There is a detailed treatment on the use of partitioned matrices for computing (i) the rank and (ii) the inverse. Specifically, if one of the diagonal blocks of a 2×2 partitioning is non-singular, then computing the rank and inverse becomes easy; the larger the non-singular block the easier to compute the rank/inverse. We can use permutation matrices to make the non-singular block as large as possible and thus facilitate computations (see example in #103).

Basic identities

- The following assume that $\mathbf{A}, \mathbf{B}$ are non-singular $n \times n$ matrices
  - $(k\mathbf{A})^{-1} = (1/k)\mathbf{A}^{-1}$
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
  - $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
  - $\text{rank}(\mathbf{AB}) = n$
- The following assume that $\mathbf{Q}, \mathbf{Q}_i$ are *orthogonal*
  - $\mathbf{QQ}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$
  - $\mathbf{Q}^{-1} = \mathbf{Q}'$
  - $\mathbf{Q}_1\mathbf{Q}_2 \ldots \mathbf{Q}_k$ is orthogonal

Important Concepts

- Left inverse (LI), right inverse (RI), inverse (*a.k.a.* left *and* right inverse)
- Orthogonal matrix: A matrix is orthogonal if its columns form an *ortho**normal*** set.
- Permutation matrices
- Schur complement of a paritioned matrix $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix}$ where $\mathbf{T}$ is non-singular is $\mathbf{Q} = \mathbf{W} - \mathbf{VT}^{-1}\mathbf{U}$. It is used, among others, when computing inverses (T8.5.11)

Notes

- **L8.1.1** is used very widely: (i) $\mathbf{A}$ has RI $\iff$ $\mathbf{A}$ has full row rank, and (ii) $\mathbf{A}$ has LI $\iff$ $\mathbf{A}$ has full col rank. (Remember: **R**ight inverse if full **r**ow rank.)
- **L8.3.2** If $\mathbf{A}$ is of full column rank, then $\mathcal{R}(\mathbf{AB}) = \mathcal{R}(\mathbf{B})$ and $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B})$. If $\mathbf{B}$ is of full row rank, then $\mathcal{C}(\mathbf{AB}) = \mathcal{C}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$.
- **C8.3.3** If $\mathbf{A}$ non-singular, then $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B})$ (i.e. rank limited by $\mathbf{B}$).
- Helmert matrix computes an orthogonal matrix from one vector whose none element is 0.
- Permutation matrices can become very convenient for computing rank (see L8.5.1 and S8.5.f)
- Block-triangular matrices are convenient for computing inverses (see L8.5.4, T8.5.11)

- The easiest is with block-diagonal matrices $\begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} \end{bmatrix}$

- Then with unit block triangular matrices, L8.5.2, $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{V} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{V} & \mathbf{I} \end{bmatrix}$ (sim. for upper up. matrix)

- Next, L8.5.4, $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{0} & \mathbf{T} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{-1} & -\mathbf{W}^{-1}\mathbf{V}\mathbf{T}^{-1} \\ \mathbf{0} & \mathbf{T}^{-1} \end{bmatrix}$ (sim. for upper low. matrix)

- Generalization (S8.5.d): The lemmas above lead to generalizations (through induction) and recursive algorithms for computing the inverses of matrices partitioned to larger than 2×2 submatrices

- **T8.5.10**: If $\mathbf{T}$ is $m \times m$ and nonsingular, then $\mathrm{rank}\begin{pmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix} = m + \mathrm{rank}(\mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U})$

- **T8.5.11**: Inverse of non-block-triangular part. m. $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix}$, given that $\mathbf{T}$ is non-singular, is computed using the results for block-triangular matrices and the Schur complement, $\mathbf{Q}$ (see def of $\mathbf{Q}$ above):

$$\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & -\mathbf{T}^{-1}\mathbf{U}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{V}\mathbf{T}^{-1} & \mathbf{Q}^{-1} \end{bmatrix} \tag{8.1}$$

- The same theorem is used to establish the non-singularity of the part. m.

- Summary on how to compute (i) the rank or (ii) the inverse of a part. m. $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$, given that either $\mathbf{A}_{11}$ or $\mathbf{A}_{22}$ is non-singular (see #101).

  - Rank: (1) Invert $\mathbf{A}_{11}$ or $\mathbf{A}_{22}$, (2) form the Schur complement, (3) find the rank of the Schur complement
  - Inverse: (1) Compute $\mathbf{A}_{11}^{-1}$ or $\mathbf{A}_{22}^{-1}$, (2) compute the Schur complement, (3) use formula (8.5.16) or (8.5.17).

Redo

- Prove L8.1.1

- Prove L8.4.1

- Prove L8.5.3

- Prove L8.5.4

- Prove T8.5.5

Chapter 9

Generalized Inverses

This chapter is about the finding inverses for matrices that are not square or nonsingular. Those so-called generalized inverses apparently become ubiquitous in statistical analysis and possibly other disciplines. The results here are used heavily in Chapter 11 where solutions to linear systems are discussed.

Important Concepts

- A *generalized inverse* (GI) of an $m \times n$ matrix $\mathbf{A}$ is any $n \times m$ matrix $\mathbf{G}$ such that $\mathbf{AGA} = \mathbf{A}$. A GI of $\mathbf{A}$ is typically denoted as $\mathbf{A}^-$

Basic Identities and Properties

- $(1/k)\mathbf{A}^-$ is a GI of $k\mathbf{A}$ (L9.3.1)

- $(\mathbf{A}^-)'$ is a GI of $\mathbf{A}'$ (L.9.3.3)

- If $\mathbf{A}$ is symmetric, then $(\mathbf{A}^-)'$ is a GI of $\mathbf{A}$ (a weaker prop compared to nonsingular $(\mathbf{A}^{-1})' = \mathbf{A}^{-1}$)

- $\text{rank}(\mathbf{A}^-\mathbf{A}) = \text{rank}(\mathbf{AA}^-) = \text{rank}(\mathbf{A})$ (C9.3.8)

- **L9.3.5**: $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A}) \iff \mathbf{B} = \mathbf{AA}^-\mathbf{B}$ and $\mathcal{R}(\mathbf{C}) \subset \mathcal{R}(\mathbf{A}) \iff \mathbf{C} = \mathbf{CA}^-\mathbf{A}$
  (A weaker prop compared to nonsingular $\mathbf{AA}^{-1} = \mathbf{I}$ or, equivalently, $\mathbf{AA}^{-1} - \mathbf{I} = \mathbf{0}$)

  - This lemma immediately relates to the consistency of a linear system:
  - **L9.5.1**: A linear system $\mathbf{AX} = \mathbf{B}$ is consistent iff $\mathbf{AA}^-\mathbf{B} = \mathbf{B}$

Notes

- Existence of GI:

  - **L9.1.4**: Every matrix has at least one GI
  - **L9.1.1**: A nonsingular matrix has exactly one GI (*i.e.* its inverse)
  - A matrix that is not nonsingular has infinite numbers of matrices (see last para of #114)

- On computing a GI of a matrix: The theorems in Sec 9.2.a give a way to compute a GI through partitioning

  - **T9.2.1**: Let $\mathbf{A}$ be $m \times n$ with rank r and partition it, $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$. Suppose $\mathbf{A}_{11}$ is nonsingular. Then, $\mathbf{G}$ is a GI of $\mathbf{A}$ iff $\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} - \mathbf{XA}_{21}\mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{Y} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{ZA}_{21}\mathbf{A}_{11}^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$ for some $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ (of appropriate dimensions)
  - The theorem above becomes particularly powerful when combined with T9.2.3
  - **L9.2.2**: Let $\mathbf{A}$ be $m \times n$ partitioned as $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ such that $\text{rank}(\mathbf{A}_{11}) = r$. Then, $\mathbf{A}_{22} = \mathbf{A}_{21}\mathbf{A}_1 1^{-1}\mathbf{A}_{12}$.
  - **T9.2.3**: Let $\mathbf{A}$ be $m \times n$ with rank r. Let $\mathbf{B}$ be obtained from $\mathbf{A}$ by permuting its rows/columns, *i.e.* $\mathbf{B} = \mathbf{PAQ}$, so that $\mathbf{B}_{11}$, the principal leading $r \times r$ submatrix of $\mathbf{B}$, is of rank r. Then, $\mathbf{G}$ is GI of $\mathbf{B}$ iff $\mathbf{G} = \mathbf{Q} \begin{bmatrix} \mathbf{B}_{11}^{-1} - \mathbf{XB}_{21}\mathbf{B}_{11}^{-1} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{Y} - \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{ZB}_{21}\mathbf{B}_{11}^{-1} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix} \mathbf{P}$

- The theorems above can be used to define a 5-step procedure to compute a GI of $\mathbf{A}$: 1) find the rank of $\mathbf{A}$, r; 2) locate the r independent rows and columns of $\mathbf{A}$; 3) Form a submatrix of $\mathbf{A}$, $\mathbf{B}_{11}$, by striking out all except those r rows and columns; 4) Compute the inverse of $\mathbf{B}_{11}$; 5) Use T9.2.1 by setting $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ to zero.

- Obtaining *every* GI in terms of a particular GI: There us an elegant theorem that provides two formulae for this purpose

  - **T9.2.7**: Let $\mathbf{A}$ be $m \times n$ and $\mathbf{G}$ a particular inverse. Then, a matrix $\mathbf{G}^*$ is a GI of $\mathbf{A}$ iff
    $\mathbf{G}^* = \mathbf{G} + \mathbf{Z} - \mathbf{GAZAG}$ for some $\mathbf{Z}$. Also, $\mathbf{G}^*$ is a GI iff
    $\mathbf{G}^* = \mathbf{G} + (\mathbf{I} - \mathbf{GA})\mathbf{T} + \mathbf{S}(\mathbf{I} - \mathbf{AG})$ for some matrices $\mathbf{T}$ and $\mathbf{S}$.
  - By letting $\mathbf{Z}$ or $\mathbf{T}, \mathbf{S}$ vary over all matrices, one can compute all GI's from one $\mathbf{G}$.

- Obtaining a matrix invariant to the choice of GI

  - **T9.4.1**: Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be matrices such that $\mathcal{R}(\mathbf{B}) \subset \mathcal{R}(\mathbf{A})$ and $\mathcal{C}(\mathbf{C}) \subset \mathcal{C}(\mathbf{B})$, then $\mathbf{BA}^-\mathbf{C}$ is invariant to the choice of $\mathbf{A}^-$. (Nearly-converse holds too.)

- The Schur treatment for GIs on Partitioned Matrices

  - A treatment similar to that of computing the inverse (see Section 8.5) applies to GIs
  - The Schur complement is now defined as $\mathbf{Q} = \mathbf{W} - \mathbf{VT}^{-1}\mathbf{U}$.
  - **T9.6.1**: Computing the rank (v similar to T8.5.10). Suppose $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ and $\mathcal{R}(\mathbf{V}) \subset \mathcal{C}(\mathbf{T})$. Then, $\text{rank} \begin{pmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix} = \text{rank}(\mathbf{T}) + \text{rank}(\mathbf{Q})$. Moreover, the GI of the partitioned matrix can be computed in terms of the GI of $\mathbf{T}^-, \mathbf{Q}^-, \mathbf{V}, \mathbf{U}$ (#121). The bottom right of the GI is $\mathbf{Q}^-$
  - **T9.6.4**: The bottom right of *any* GI of the above partitioned matrix is a GI of $\mathbf{Q}$.

- Comparison of the 5-step algorithm (above) and the Schur way: 5-step requires finding a basis that spans $\mathbf{A}$, whereas Schur-way requires $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ and $\mathcal{R}(\mathbf{V}) \subset \mathcal{C}(\mathbf{T})$.

Important Lemmas and Theorems

- **L9.3.5**: See about $\mathbf{B} = \mathbf{AA}^-\mathbf{B}$ – see definition above in Basic Identities and Properties

- **T9.2.7**: On computing all GIs from a particular GI (above)

- **T9.4.1**: Obtaining an matrix invariant to the choice of GI

Redo

- Prove T9.1.2

- Prove T9.2.7

- Prove T9.4.1

- Prove T9.6.4

- Solve Exercise 7

- Solve Exercise 12a

Chapter 10

Idempotent Matrices

Apparently the class of idempotent matrices is very important in statistics and other disciplines. They do have some very interesting and non-obvious properties (*e.g.* L10.1.1, L10.1.2 or T10.2.1)

Definition and Some Properties

- $\mathbf{A}$ is said to be idempotent if $\mathbf{A}^2 = A$

- The only $n \times n$ idempotent matrix is $\mathbf{I}_n$ (L10.1.1)

- $\mathbf{A}'$ is idempotent $\iff$ $\mathbf{A}$ is idempotent (L10.1.2)

- $\mathbf{A}$ is idempotent $\iff$ $\mathbf{I} - \mathbf{A}$ is idempotent (L10.1.2)

Notes

- One of the most non-intuitive relations is through the following theorem (and its corollary):

    - **T10.2.1**: For any square matrix $\mathbf{A}$ such that $\mathbf{A}^2 = k\mathbf{A}$, $\mathrm{tr}(\mathbf{A}) = k\,\mathrm{rank}(\mathbf{A})$
    - **C10.2.2**: For any idempotent matrix $\mathbf{A}$, $\mathrm{rank}(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$ (I guess I find this non-intuitive because the rank –an integer– is computed with tr which does not in general yield an integer!)
    - **T10.2.7**: $\mathbf{B}$ is GI of $\mathbf{A}$ iff $\mathbf{BA}$ is idempotent and $\mathrm{rank}(\mathbf{BA}) = \mathrm{rank}(\mathbf{A})$ (This also is non-intuitive b/c idempotent matrices seem like they are rare but yet so easy to obtain from a matrix and its inverse!)
    - **L10.2.4**: For any idempotent matrix, $\mathrm{rank}(\mathbf{I} - \mathbf{A}) = \mathrm{tr}(\mathbf{I} - \mathbf{A}) = n - \mathrm{rank}(\mathbf{A})$.

Redo

- Prove T10.2.1

- Prove T10.2.7

- Solve Exercise 3

Chapter 11

Linear Systems: Solutions

Notes

- Relationships between $\mathcal{N}(\mathbf{A})$ and $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{X})$
    - $\mathcal{N}(\mathbf{A}) \subset \mathcal{C}(\mathbf{I} - \mathbf{A})$
    - $\mathcal{N}(\mathbf{I} - \mathbf{A}) \subset \mathcal{C}(\mathbf{A})$ (p140)
    - **C11.2.2**: $\mathcal{N}(\mathbf{A}) = \mathcal{C}(\mathbf{I} - \mathbf{A}^-\mathbf{A})$
    - (1) If $\mathbf{AX} = \mathbf{0}$, then $\mathcal{C}(\mathbf{X}) \subset \mathcal{N}(\mathbf{A})$. (2) If also rank$(\mathbf{X}) = n - \text{rank}(\mathbf{A})$, then $\mathcal{C}(\mathbf{X}) = \mathcal{N}(\mathbf{A})$
    - **T11.7.1** Let $\mathbf{A}$ be $n \times n$. Then, $\mathcal{N}(\mathbf{A}) = \mathcal{C}(\mathbf{I} - \mathbf{A}) \iff \mathbf{A}$ is idempotent.

- General forms of solutions (useful for finding/checking alternative solutions etc.)
    - **T11.2.1** (Homogeneous): $\mathbf{X}^*$ is solution to homogeneous LS $\mathbf{AX} = \mathbf{0} \iff \mathbf{X}^* = (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{Y}$ for some $\mathbf{Y}$
    - **T11.2.3** (Nonhom.): $\mathbf{X}_0$ is solution to $\mathbf{AX} = \mathbf{B}$. $\mathbf{X}^*$ is sol. $\iff \mathbf{X}^* = \mathbf{X}_0 + \mathbf{Z}^*$ for some $\mathbf{Z}^*$ sol. to $\mathbf{AZ} = \mathbf{0}$
    - **T11.2.4** (Nonhom.): $\mathbf{X}^*$ is solution to $\mathbf{AX} = \mathbf{B} \iff \mathbf{X}^* = \mathbf{A}^-\mathbf{B} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{Y}$ for some $\mathbf{Y}$
    - **T11.5.1**: Let rank$(\mathbf{A}) = n$, rank$(\mathbf{B}) = p$. Then, $\mathbf{X}^*$ is solution to $\mathbf{AX} = \mathbf{B} \iff \mathbf{X}^* = \mathbf{A}^-\mathbf{B}$ for some GI $\mathbf{A}^-$

- Number of solutions
    - **L11.3.1** (homogeneous, vec.): $\mathbf{A}$ is $m \times n$. Then, $\dim[\mathcal{N}(\mathbf{A})] = n - \text{rank}(\mathbf{A})$. Note: system is $\mathbf{Ax} = \mathbf{0}$
    - **L11.3.2** (homogeneous, mat.) The dimension of solution space for $\mathbf{AX} = \mathbf{0}$ (in $n \times p$ matrix $\mathbf{X}$) is $p[n - \text{rank}(\mathbf{A})]$
    - Summary based on p142-143:
        * (homog.) If $\mathbf{A}$ has full col. rank, $\mathbf{AX} = \mathbf{0}$ has *one* solution (*i.e.* the null mat. $\mathbf{0}$)
        * (homog.) If $\mathbf{A}$ is col. rank deficient, then $\mathbf{AX} = 0$ has infinite solutions
        * (nonhom.) Since there is one-to-one correspondence between solutions to $\mathbf{AX} = \mathbf{0}$ and sol.s to $\mathbf{AX} = \mathbf{B}$, the two above apply to nonhomogeneous case too given that the nonhom. system is consistent

- Equivalence of linear systems: When are the solution sets of two linear systems the same?
    - **L11.6.1.** If $\mathbf{C}$ is of full col. rank, then the systems $\mathbf{AX} = \mathbf{B}$ and $\mathbf{CAX} = \mathbf{CB}$ are equivalent
    - **L11.6.2.** For any $\mathbf{A}$, the linear systems $\mathbf{A'AX} = \mathbf{A'AF}$ and $\mathbf{AX} = \mathbf{AF}$ are equivalent
    - **Exercise 6.** If rank$(\mathbf{CD}) = \text{rank}(\mathbf{D})$, then system $\mathbf{CDAX} = \mathbf{CDB}$ is equivalent to the system $\mathbf{DAX} = \mathbf{DB}$
    - **Exercise 7.** If rank$[\mathbf{C}(\mathbf{A}, \mathbf{B})] = \text{rank}(\mathbf{A}, \mathbf{B})$, then the systems $\mathbf{AX} = \mathbf{B}$ and $\mathbf{CAX} = \mathbf{CB}$ are equivalent

- Can we derive a vector/matrix $\mathbf{K'X}$ invariant to the choice of solution $\mathbf{X}$?
    - **T11.10.1.** Let $\mathbf{K}$ be $n \times q$. Then, $\mathbf{K'X}$ is invariant to solution $\mathbf{X} \iff \mathcal{R}(\mathbf{K'}) \subset \mathcal{R}(\mathbf{A})$
    - **T11.10.3.** Let $\mathbf{X}, \mathbf{Y}$ be solutions to $\mathbf{AX} = \mathbf{B}$ and $\mathbf{A'Y} = \mathbf{K}$. Then, $\mathbf{K'X}_0 = \mathbf{Y'}_0\mathbf{B}$ for any solutions $\mathbf{X}_0, \mathbf{Y}_0$.

- Approaches to solve linear systems
    - If $\mathbf{A}$ is upper block-triangular such that the diagonal blocks $\mathbf{A}_{ii}$ are non-singular, then back-substitution (Section 11.8)
    - If $\mathbf{A}$ is lower block-triangular such that the diagonal blocks $\mathbf{A}_{ii}$ are non-singular, then forward elimination
    - Find a decomposition $\mathbf{A} = \mathbf{KT}$ such that solving becomes easier. More on Ch. 14 and 21 (p.149)
    - **T11.11.1.** Absorption: see Section p.152 and 4-step summary in p. 153

Chapter 12

Projections and Projection Matrices

Definitions and some properties

- *Orthogonality*: If matrix $\mathbf{Y}$ is orthogonal to every matrix in a subspace $\mathcal{U}$, then $\mathbf{Y}$ is orthogonal to $\mathcal{U}$: $\mathbf{Y} \perp \mathcal{U}$.

- *Projection*: Let $\mathbf{Y} \in \mathcal{V}$, $\mathbf{Z} \in \mathcal{V}$ and $\mathcal{U} \subset \mathcal{V}$. $\mathbf{Z}$ is the projection of $\mathbf{Y}$ on $\mathcal{U}$ if $(\mathbf{Y} - \mathbf{Z}) \perp \mathcal{U}$
  For any $\mathbf{Y} \in \mathcal{U}$, $\mathbf{Y}$ itself is the projection of $\mathbf{Y}$ on $\mathcal{U}$

- The *orthogonal complement* $\mathcal{U}^\perp$ of $\mathcal{U} \subset \mathcal{V}$ is the set of all matrices in $\mathcal{V}$ that are orthogonal to $\mathcal{U}$
  - $\mathcal{U} = (\mathcal{U}^\perp)^\perp$ (T12.5.4)
  - $\mathcal{W} \subset \mathcal{U} \iff \mathcal{U}^\perp \subset \mathcal{W}^\perp$ (C12.5.6)
  - $\dim(\mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{U}^\perp)$ for $\mathcal{U} \subset \mathcal{U}^\perp$ (T12.5.12)

- *Projection matrix*: $\mathbf{P_X} := \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$. That is, $\mathbf{P_X}\mathbf{y}$ is the projection of any $\mathbf{y} \in \mathcal{R}^n$ on $\mathcal{C}(\mathbf{X})$
  - $\mathbf{P_X}\mathbf{X} = \mathbf{X}$ (T12.3.4)
  - $\mathbf{P_X}' = \mathbf{P_X}$ (*i.e.* symmetric, T12.3.4)
  - $\mathbf{P_X}^2 = \mathbf{P_X}$ (*i.e.* idempotent, T12.3.4)
  - $(\mathbf{I} - \mathbf{P_X})$ is also idempotent and symmetric
  - $(\mathbf{I} - \mathbf{P_X})$ is the projection matrix for $\mathcal{C}^\perp(\mathbf{X})$ (C12.5.10)

Notes

- Orthogonality conditions
  - **L12.1.1**. $\mathbf{Y} \perp \mathcal{U} = \mathrm{sp}(\mathbf{X}_1, \ldots, \mathbf{X}_k) \iff \mathbf{Y} \cdot \mathbf{X}_i$ for $i = 1, ..., k$. $\mathcal{U} \perp \mathcal{V} = \mathrm{sp}(\mathbf{Z}_1, \ldots, \mathbf{Z}_t) \iff \mathbf{X}_i \perp \mathbf{Z}_j \ \forall i, j$.
  - **C12.1.2**. $\mathbf{y} \, (\in \mathcal{R}^m) \perp \mathcal{C}(\mathbf{X}) \iff \mathbf{X}'\mathbf{y} = 0$
  - **T12.1.3**. Let $\mathbf{Y} \in \mathcal{V}$, $\mathcal{U} \subset \mathcal{V}$ and $\dim(\mathcal{U}) = r$. Then, there is *unique* $\mathbf{Z} \in \mathcal{U}$ s.t. $(\mathbf{Y} - \mathbf{Z}) \perp \mathcal{U}$. Let $\{\mathbf{X}_1, \ldots, \mathbf{X}_r\}$ be an orthonormal basis for $\mathcal{U}$. Then, $\mathbf{Z} = c_1\mathbf{X}_1 + \ldots + c_r\mathbf{X}_r$ where $c_i = \mathbf{X} \cdot \mathbf{Y}_i$. Moreover, $\mathbf{Z} = \mathbf{Y} \iff \mathbf{Y} \in \mathcal{U}$.

- Projection of column vector and Projection Matrix
  - **T12.2.1**. $\mathbf{z}$ is the projection of $\mathbf{y} \in \mathcal{R}^n$ on $\mathcal{U} \subset \mathcal{R}^n \implies \mathbf{z} = \mathbf{X}\mathbf{b}^*$ for any solution $\mathbf{b}^*$ to $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$
    Also (C12.2.2), $\mathbf{z} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$
  - **T12.3.1** (uniqueness). Let $\mathcal{U} \subset \mathcal{R}^n$. There is *unique* $n \times n$ matrix $\mathbf{A}$ s.t. $\mathbf{A}\mathbf{y}$ is the projection of any $\mathbf{y} \in \mathcal{R}^n$ on $\mathcal{U}$. For any $\mathbf{X}$ s.t. $\mathcal{C}(\mathbf{X}) = \mathcal{U}$, $\mathbf{A} = \mathbf{P_X}$. (see def. of $\mathbf{P_X}$ above)
  - **T12.3.4** (other outcomes of this theorem are properties above). $\mathbf{P_X} = \mathbf{X}\mathbf{B}^*$ for any solution to $\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'$
  - **T12.5.8** (orth. compl). Let $\mathbf{Y} \in \mathcal{V}$, $\mathcal{U} \subset \mathcal{V}$. Projection of $\mathbf{Y}$ on $\mathcal{U}^\perp$ is $\mathbf{Y} - \mathbf{Z}$ where $\mathbf{Z}$ is proj. of $\mathbf{Y}$ on $\mathcal{U}$
  - **C12.5.10**. $(\mathbf{I} - \mathbf{P_X})$ is the projection matrix for $\mathcal{C}^\perp(\mathbf{X})$
  - **T12.5.11**. Let $\mathbf{Y} \in \mathcal{V}$, $\mathcal{U} \subset \mathcal{V}$. $\exists$ *unique* $\mathbf{Z} \in \mathcal{U}$, $\mathbf{W} \in \mathcal{U}^\perp$ s.t. $\mathbf{Y} = \mathbf{Z} + \mathbf{W}$. Also $\mathbf{Z}$ ($\mathbf{W}$) is proj. of $\mathbf{Y}$ on $\mathcal{U}$ ($\mathcal{U}^\perp$).

- Least squares connection
  - **T12.4.1**. Let $\mathbf{Y} \in \mathcal{V}$ and $\mathcal{U} \subset V$. Then, for $\mathbf{W} \in \mathcal{U}$, $||\mathbf{Y} - \mathbf{W}||$ is minimized by $\mathbf{W} = \mathbf{Z}$ where $\mathbf{Z}$ is $\mathbf{Y}$'s projection on $\mathcal{U}$. Moreover, $||\mathbf{Y} - \mathbf{Z}||^2 = \mathbf{Y} \cdot (\mathbf{Y} - \mathbf{Z})$.
  - **T12.4.2** (Residue). For a column vector projection $\mathbf{z}$, the residue $||\mathbf{y} - \mathbf{z}||^2$ is $\mathbf{y}'(\mathbf{I} - \mathbf{P_X})\mathbf{y}$

Chapter 13

Determinant

13.1 Definitions and some properties

- *Determinant* of $\mathbf{A}$: $\det(\mathbf{A}) = |\mathbf{A}| = \sum_{\mathbf{j} \in \mathcal{J}} (-1)^{\phi_n(j_1, \ldots, j_n)} a_{1j_1} \ldots a_{nj_n}$ where $\mathbf{j} = (j_1, \ldots, j_n)$ and $\mathcal{J}$ is the set that contains all permutations of first n positive integers, and $\phi(\cdot)$ is, *e.g.* $\phi_5(3, 7, 2, 1, 4) = 2 + 3 + 1 + 0 = 6$ (p178)

  - $|\mathbf{A}'| = |\mathbf{A}|$ (L13.2.1)
  - If $\mathbf{B}$ is same with $\mathbf{A}$ except that one of its rows/columns is multiplied with k, then $|\mathbf{B}| = k|\mathbf{A}|$ (L13.2.1)
  - $|k\mathbf{A}| = k^n |\mathbf{A}|$ (C13.2.4)
  - $|-\mathbf{A}| = (-1)^n |\mathbf{A}|$
  - If $\mathbf{B}$ is formed from $\mathbf{A}$ by interchanging 2 of its columns or rows, then $|\mathbf{B}| = -|\mathbf{A}|$ (T13.2.6)
  - If $\mathbf{B}$ is $n \times p$ and $\mathbf{C}$ is $n \times q$, then $|\mathbf{B}, \mathbf{C}| = (-1)^{pq}|\mathbf{B}, \mathbf{C}|$ (T13.2.7)
  - $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$, and more generally $|\mathbf{A}_{11}\mathbf{A}_{22} \ldots \mathbf{A}_{rr}| = |\mathbf{A}_{11}||\mathbf{A}_{22}| \ldots |\mathbf{A}_{rr}|$ (T13.3.4)
  - $|\mathbf{A}^k| = |\mathbf{A}|^k$
  - $|\mathbf{A}'\mathbf{A}| = |\mathbf{A}|^2$
  - $\mathbf{A}$ is nonsingular $\iff |\mathbf{A}| \neq 0$ (T13.3.7)
  - If $\mathbf{A}$ is nonsingular, then $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$

- Cofactor, minor and adjoint

  - *Minor.* Let $\mathbf{A}_{ij}$ be the submatrix of $\mathbf{A}$ obtained by striking out its ith row and jth column. Then, $|\mathbf{A}_{ij}|$ is called the minor of the element a_{ij} of $\mathbf{A}$.
  - *Cofactor*, α_{ij} is the signed minor: $\alpha_{ij} = (-1)^{i+j}|\mathbf{A}_{ij}|$
  - The *adjoint* of $\mathbf{A}$ is the (transposed) matrix of cofactors: $\mathrm{adj}(\mathbf{A}) = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{n1} \\ \vdots & \ddots & \vdots \\ \alpha_{1n} & \ldots & \alpha_{nn} \end{bmatrix}$

- Important properties of adjoint

  - $\mathbf{A}\,\mathrm{adj}(\mathbf{A}) = |\mathbf{A}|\mathbf{I}_n$ (T13.5.3)
  - If $\mathbf{A}$ is nonsingular, then $\mathbf{A}^{-1} = (1/|\mathbf{A}|)\,\mathrm{adj}(\mathbf{A})$
  - $\mathrm{adj}(\mathbf{A}\mathbf{B}) = \mathrm{adj}(\mathbf{B})\mathrm{adj}(\mathbf{A})$ (Exercise 14)

13.2 Notes

- Computation in special cases

  - If $\mathbf{A}$ is (upper or lower) triangular, then $|\mathbf{A}| = a_{11}a_{22} \ldots a_{nn}$
  - **T13.1.1**. $\begin{vmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{V} & \mathbf{W} \end{vmatrix} = |\mathbf{T}||\mathbf{W}|$. Repeated application $\implies$ For any block triangular $\mathbf{A}$, $|\mathbf{A}| = |\mathbf{A}_{11}||\mathbf{A}_{22}| \ldots |\mathbf{A}_{rr}|$.
  - **C13.3.2**. $\begin{vmatrix} \mathbf{0} & \mathbf{T} \\ \mathbf{W} & \mathbf{V} \end{vmatrix} = \begin{vmatrix} \mathbf{V} & \mathbf{W} \\ \mathbf{T} & \mathbf{0} \end{vmatrix} = (-1)^{mn}|\mathbf{T}||\mathbf{W}|$

- If $\mathbf{T}$ is nonsingular, then $\begin{vmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{vmatrix} = |\mathbf{T}||\mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}|$

- Some cases where determinants are the same

  - **T13.2.10**. If $\mathbf{B}$ is formed from $\mathbf{A}$ by adding to any one row or column of $\mathbf{A}$, scalar multiples of one or more other rows, then $|\mathbf{B}| = |\mathbf{A}|$.
  - **T13.2.11**. For any upper/lower *unit* triangular $\mathbf{T}$, $|\mathbf{A}\mathbf{T}| = |\mathbf{T}\mathbf{A}| = |\mathbf{A}|$

- Some cases where determinant is 0 or ± 1

  - **L13.2.9**. If a row or column of $\mathbf{A}$ is multiple of another row or column, then $|\mathbf{A}| = 0$
  - **T13.3.7**. $|\mathbf{A}| = 0 \iff \mathbf{A}$ is nonsingular (see also properties above)
  - **C13.3.6** For any orthogonal matrix $\mathbf{P}$, $|\mathbf{P}| = \pm 1$

Chapter 14

Linear, Bilinear and Quadratic Forms

14.1 Definitions and some properties

- Linear, bilinear, quadratic forms

  - Linear form: $\mathbf{ax}$, linear function: $f(\mathbf{x})$. $\mathbf{a}$: coefficient vector of linear form
  - Bilinear form: $\mathbf{x}'\mathbf{Ay}$, bilinear function: $f(\mathbf{x}, \mathbf{y})$. $\mathbf{A}$: matrix of the linear form
  - Symmetric bilinear form is when $\mathbf{x}'\mathbf{Ay} = \mathbf{y}'\mathbf{Ax}$ for all $\mathbf{x}, \mathbf{y}$.
  - Quadratic form: $\mathbf{x}'\mathbf{Ax}$. $\mathbf{A}$: matrix of the quadratic form.
  - An $n \times n$ matrix is skew-symmetrix if $\mathbf{A}' = -\mathbf{A}$ (which implies $a_{ii} = 0$ for $i = 1, \ldots, n$).

- Nonnegative definite quadratic forms

  - A quadratic form $\mathbf{x}'\mathbf{Ax}$ (or its matrix $\mathbf{A}$) is called *nonnegative definite* if $\mathbf{x}'\mathbf{Ax} \geq 0$ for all $\mathbf{x} \in \mathcal{R}^n$
  - A quadratic form $\mathbf{x}'\mathbf{Ax}$ (or its matrix $\mathbf{A}$) is called *positive definite* if $\mathbf{x}'\mathbf{Ax} > 0$ for all $\mathbf{x} \in \mathcal{R}^n$
  - A nonnegative quadratic form $\mathbf{x}'\mathbf{Ax}$ (or its matrix $\mathbf{A}$) is called *positive semidefinite* if there exist nonnull $\mathbf{x} \in \mathcal{R}^n$ such that $\mathbf{x}'\mathbf{Ax} = 0$.
  - A quadratic form $\mathbf{x}'\mathbf{Ax}$ (or its matrix $\mathbf{A}$) is called *nonpositive definite*, *negative definite* or *negative semidefinite* if $-\mathbf{x}'\mathbf{Ax}$ is nonnegative definite, positive definite or positive semidefinite.
  - A quadratic form that is neither nonnegative nor nonpositive definite is called *indefinite*.

- Some properties of nonnegative matrices/forms

  - If $\mathbf{A}$ is nonnegative and k is a scalar such that $k > 0$, then $k\mathbf{A}$ is also nonnegative (L14.2.3).
  - If $\mathbf{A}, \mathbf{B}$ are nonnegative, then $\mathbf{A} + \mathbf{B}$ is nonnegative. If one of them is positive, then $\mathbf{A} + \mathbf{B}$ is positive. (L14.2.4)
  - Any positive definite matrix is nonsingular (L14.2.8).
  - The inverse of a positive definite (semidefinite nonsingular) matrix is positive definite (semidefinite). (C13.2.11)
  - Every symmetric idempotent matrix is nonnegative definite (L14.2.7).
  - A symmetric nonnegative definite $\mathbf{A}$ is positive definite $\iff$ it $\mathbf{A}$ is nonsingular (C14.3.12).

- Properties regarding traces of nonnegative matrices

  - $\operatorname{tr}(\mathbf{A}) > 0$ for any positive definite $\mathbf{A}$. (T14.7.1)
  - $\operatorname{tr}(\mathbf{A}) \geq 0$ for any nonnegative matrix, with equality holding iff $\mathbf{A}$ is skew-symmetric. (T14.7.2)
  - For any positive $\mathbf{A}$ and nonnull symmetric nonnegative $\mathbf{B}$, $\operatorname{tr}(\mathbf{AB}) > 0$. (T14.7.4)
  - For any positive $\mathbf{A}$ and symmetric nonnegative $\mathbf{B}$, $\operatorname{tr}(\mathbf{AB}) \geq 0$ with equality holding iff $\mathbf{B} = \mathbf{0}$. (C14.7.5)
  - For any nonnegative $\mathbf{A}$ and symmetric nonneg. $\mathbf{B}$, $\operatorname{tr}(\mathbf{AB}) \geq 0$; equality holds iff $\mathbf{BAB}$ is skew-symm. (T14.7.6)

- Properties of determinants of nonnegative matrices

  - For any symmetric positive $\mathbf{A}$, $|\mathbf{A}| > 0$; for any symmetric positive semidefinite $\mathbf{A}$, $|\mathbf{A}| = 0$ (T14.9.1)
  - For nonnegative $\mathbf{A}$, $|\mathbf{A}| \geq 0$. For positive definite or nonsingular nonsymmetric positive semidefinite $\mathbf{A}$, $|\mathbf{A}| > 0$. For symmetric positive semidefinite or singular nonsymmetric positive semidefinite $\mathbf{A}$, $|\mathbf{A}| = 0$ (T14.9.3, C14.9.4)

- Inner products, orthogonality, projections, orthogonal complements

- *Inner product* bw $\mathbf{x}, \mathbf{y}$ wrt a spd. (symmetric and positive definite) matrix $\mathbf{W}$ is $\mathbf{x}'\mathbf{W}\mathbf{y}$. (see also below)

- *Orthogonality.* $\mathbf{x} \perp_{\mathbf{W}} \mathbf{y} \iff \mathbf{x}'\mathbf{W}\mathbf{y} = 0$ for spd. $\mathbf{W}$.

- $\mathbf{x} \perp_{\mathbf{W}} \mathcal{U} \iff \mathbf{x}'\mathbf{W}\mathbf{y} = 0 \; \forall \mathbf{y} \in \mathcal{U}$

- $\mathcal{U} \perp_{\mathbf{W}} \mathcal{V} \iff \mathbf{x}'\mathbf{W}\mathbf{y} = 0 \; \forall \mathbf{x} \in \mathcal{U}, \mathbf{y} \in \mathcal{V}$

- *Projection matrix* wrt spd. $\mathbf{W}$ is $\mathbf{P}_{\mathbf{X},\mathbf{W}} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}$

- *Projection* of $\mathbf{y} \in \mathcal{R}^n$ to $\mathcal{U} \subset \mathcal{R}^n$ wrt spd. $\mathbf{W}$ is $\mathbf{z} \in \mathcal{U}$ such that $\mathbf{y} - \mathbf{z} \perp_{\mathbf{W}} \mathcal{U}$.

- *Generalized or weighted least squares problem* is minimizing $(\mathbf{y} - \mathbf{u})'\mathbf{W}(\mathbf{y} - \mathbf{u})$ for $\mathbf{u}$ wrt spd. $\mathbf{W}$.

- *Orthogonal complement* of $\mathcal{C}(\mathbf{X})$ is $\mathcal{C}_{\mathbf{W}}^{\perp}(\mathbf{X}) = \{\mathbf{y} \in \mathcal{R}^n : \mathbf{X}'\mathbf{W}\mathbf{y} = 0\} = \mathcal{N}(X'W) = \mathcal{C}(\mathbf{I} - \mathbf{P}_{\mathbf{X},\mathbf{w}})$

- *Gram matrix* (or *Gramian*). Let $\mathbf{X}_1, \ldots, \mathbf{X}_r \in \mathcal{V}$. Gramian is the matrix whose ijth element is $\mathbf{X}_i \cdot \mathbf{X}_j$ (*i.e.* inner product).

14.2 Decompositions of matrices and quadratic forms

This section presents some fundamental decompositions. There is a specific focus on the decomposition of symmetric matrices and positive or nonnegative definite matrices, but there are generic decompositions as well.

Any square matrices. Let $\mathbf{A}$ be a square matrix.

- $\exists$ nonsingular $\mathbf{P}$ and n scalars $d_1, \ldots, d_n$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} d_i y_i^2$ for $\mathbf{y} = \mathbf{P}\mathbf{x}$ (C14.3.6)

- **LDU** decomposition (**L** (**U**): lower (upper) unit triangular, **D**: diagonal). Let $\mathbf{A}_i$ be the ith order leading principal submatrix of $\mathbf{A}$, and partition it as
$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i-1} & \mathbf{a}_i \\ \mathbf{b}_i' & a_{ii} \end{bmatrix}.$$
  $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$ if for $i = 2, \ldots, n$, $\mathbf{a}_i \in \mathcal{C}(\mathbf{A}_{i-1})$ and $\mathbf{b}_i' \in \mathcal{R}(\mathbf{A}_{i-1})$. (T14.5.4)

For symmetric matrices. Let $\mathbf{A}$ be a symmetric matrix.

- $\exists$ nonsingular $\mathbf{Q}$ s.t. $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ is a diagonal matrix. (T14.3.4)

- $\exists$ nonsingular $\mathbf{P}$ and diagonal $\mathbf{D}$ s.t. $\mathbf{A} = \mathbf{P}'\mathbf{D}\mathbf{P}$. (C14.3.5)

- $\mathbf{A}$ has $\mathbf{U}'\mathbf{D}\mathbf{U}$ decomposition if $\mathbf{a}_i \in \mathcal{C}(\mathbf{A}_{i-1})$ (see LDU decomposition above for notation) (T14.5.4)

For symmetric nonnegative matrices. Let $\mathbf{A}$ be an $n \times n$ symmetric nonnegative matrix.

- $\exists \mathbf{P}$ s.t. $\mathbf{A} = \mathbf{P}'\mathbf{P}$. (T14.3.7, C14.3.8 and C14.3.10)

- $\mathbf{A}$ has a U'DU decomposition (with nonnegative diagonal elements). (C14.5.15)

- Let $r = \text{rank}(\mathbf{A})$. There is unique upper triangular $\mathbf{T}$ with $n - r$ null rows s.t. $\mathbf{A} = \mathbf{T}'\mathbf{T}$.

- Let $\mathbf{A} = \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$. Then, $\exists \mathbf{R}, \mathbf{S}$ such that $\mathbf{T} = \mathbf{R}'\mathbf{R}$, $\mathbf{U} = \mathbf{R}'\mathbf{S}$, $\mathbf{V} = \mathbf{S}'\mathbf{R}$, $\mathbf{W} = \mathbf{S}'\mathbf{S}$.

For positive definite matrices. Let $\mathbf{A}$ be a positive definite matrix.

- $\mathbf{A}$ has a unique LDU decomposition (T14.5.9).

For positive and symmetric matrices. Let $\mathbf{A}$ be a positive definite and symmetric matrix.

- $\exists$ nonsingular $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}'\mathbf{P}$ (C14.3.13).

- $\mathbf{A}$ has a unique U'DU decomposition (C14.5.10).

- *Cholesky decomposition.* $\exists$ unique upper triangular $\mathbf{T}$ s.t. $\mathbf{A} = \mathbf{T}'\mathbf{T}$. (T14.5.11)

14.3 Notes

- Equivalence/symmetry of bilinear/quadratic forms

  - A bilinear form $\mathbf{x}'\mathbf{A}\mathbf{y}$ is symmetric $\iff$ $\mathbf{A}$ is a symmetric matrix (p209).
  - **L14.1.1**. Two quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$ $\mathbf{x}'\mathbf{B}\mathbf{x}$ are identically equal $\iff$ $\mathbf{A} + \mathbf{A}' = \mathbf{B} + \mathbf{B}'$
  - **C14.1.2** For any quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$, there is a *unique* symmetric matrix $\mathbf{B}$ such as $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$, namely $\mathbf{B} = (1/2)(\mathbf{A} + \mathbf{A}')$

- On nonnegative and positive definiteness

  - **T14.2.9**. Let $\mathbf{A}$ be $n \times n$ and $\mathbf{P}$ any $n \times m$. (1) If $\mathbf{A}$ is nonnegative, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is nonnegative. (2) If $\mathbf{A}$ nonnegative and rank$(\mathbf{P}) < m$, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive semidefinite. (3) If $\mathbf{A}$ is positive, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive.
  - **C14.2.10**. If $\mathbf{P}$ is nonsingular and $\mathbf{A}$ pos. definite (semidefinite), then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is pos. definite (semidefinite).
  - **C14.2.12**. Any principal submatrix of a positive definite (semidefinite) matrix is positive (nonnegative) definite.
  - **C14.2.13**. The diagonal elements of a positive (nonnegative) definite matrix are positive (nonnegative).
  - **C14.2.14**. Let $\mathbf{P}$ be any $n \times m$. If rank$(\mathbf{P}) = m$, $\mathbf{P}'\mathbf{P}$ is positive definite, otherwise, it's positive semidefinite.
  - **C14.2.14**. Let, $\mathbf{P}$ be nonsingular and $\mathbf{D}$ diagonal. Then, (1) $\mathbf{P}'\mathbf{D}\mathbf{P}$ is nonnegative (positive) $\iff$ $\mathbf{D}$'s elements are nonnegative (positive). (2) $\mathbf{P}'\mathbf{D}\mathbf{P}$ is positive definite if $\mathbf{D}$'s elements are nonnegative and include some zeros.

- Decomposition of symmetric matrices

  - **T14.3.4**. For any $n \times n$ symmetric $\mathbf{A}$, $\exists$ nonsingular $\mathbf{Q}$ such that $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ is a diagonal matrix.
  - **T14.3.4**. For any $n \times n$ symmetric $\mathbf{A}$, $\exists$ nonsingular $\mathbf{P}$ and diagonal $\mathbf{D}$ such that $\mathbf{A} = \mathbf{P}'\mathbf{D}\mathbf{P}$.
  - **C14.3.6**. For any $n \times n$ $\mathbf{A}$, $\exists$ nonsingular $\mathbf{P}$ and scalars $d_1, \ldots, d_n$ s.t. the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ is expressible as a linear combination $\sum_{i=1}^{n} d_i y_i^2$ of the squares of the elements $y_1, \ldots, y_n$ of the transformed vector $\mathbf{y} = \mathbf{P}\mathbf{x}$

- Decomposition of symmetric nonnegative matrices

  - **T14.3.7**. A non-null $n \times n$ $\mathbf{A}$ is symmetric nonnegative of rank r $\iff$ $\exists$ $r \times n$ $\mathbf{A}$ s.t. $\mathbf{A} = \mathbf{P}'\mathbf{P}$.
  - **C14.3.9** (and **C14.3.13**). Let rank$(\mathbf{A} + \mathbf{A}') = r$. The quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ is nonnegative (or positive assuming $r = n$) $\iff$ $\exists$ $r \times n$ $\mathbf{P}$ s.t. $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{x}'\mathbf{P}\mathbf{x}$ $\forall \mathbf{x}$.
  - **C14.3.13**. An $n \times n$ symmetric $\mathbf{A}$ is positive definite $\iff$ $\exists$ nonsingular $\mathbf{P}$ s.t. $\mathbf{A} = \mathbf{P}'\mathbf{P}$.
  - **L14.4.1**. Every symmetric nonnegative matrix has a symmetric positive definite generalized inverse.

- Some results regarding the uniqueness of LDU decomposition

  - (**T14.5.6**) *If* $\mathbf{A}$ has an LDU decomposition and if $\mathbf{A}$'s $n - 1$ order principal submatrix is nonsingular, then the LDU decomposition is unique.
  - (**C14.5.7**) $\mathbf{A}$ has unique LDU decomposition if its ith order principal submatrix is nonsingular for $i = 1, \ldots, n-1$
  - (**L14.5.8**) If a symmetric $\mathbf{A}$ has a unique LDU decomposition, then it's actually an U'DU decomposition.

- Obtaining generalized inverse from LDU or U'DU decomposition.

  - (p236). If $\mathbf{A}$ has LDU (U'DU) decomposition, then $\mathbf{G} = \mathbf{U}^{-1}\mathbf{D}^{-}\mathbf{L}^{-1}$ $[\mathbf{G} = \mathbf{U}^{-1}\mathbf{D}^{-}(\mathbf{U}^{-1})']$ is a GI.
  - (p236). Assume $\mathbf{A}$ has Cholesky dec. $\mathbf{A} = \mathbf{T}'\mathbf{T}$ and let T_1 be the submatrix of $\mathbf{T}$ obtained by striking out its null rows, and let R be the right inverse of $\mathbf{T}_1$. Then, $\mathbf{R}\mathbf{R}'$ is a GI of $\mathbf{A}$.

- On block-diagonal and partitioned matrices

  - **C14.8.2**. If $\mathbf{A} = \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$ is sym. nonneg., then, $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T}), \mathcal{R}(\mathbf{V}) \subset \mathcal{R}(\mathbf{T}), \mathcal{C}(\mathbf{V}) \subset \mathcal{C}(\mathbf{W}), \mathcal{R}(\mathbf{U}) \subset \mathcal{R}(\mathbf{W})$.
  - **L14.8.3**. Let $\mathbf{A} = \text{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_k)$. Then, $\mathbf{A}$ is nonneg (positive) $\iff$ $\mathbf{A}_1, \ldots, \mathbf{A}_k$ are nonneg (positive).
  - **T14.8.4**. Let $\mathbf{A} = \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$. (1) $\mathbf{A}$ is positive definite $\implies$ Schur complements $\mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}$ and $\mathbf{T} - \mathbf{U}\mathbf{W}^{-1}\mathbf{V}$ are positive. (2) $\mathbf{A}$ is symmetric positive semidefinite $\implies$ Schur complements $\mathbf{W} - \mathbf{V}\mathbf{T}^{-}\mathbf{U}$ and $\mathbf{T} - \mathbf{U}\mathbf{W}^{-}\mathbf{V}$ are nonnegative. (3) $\mathbf{A}$ is positive semidefinite and $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ or $\mathcal{R}(\mathbf{V}) \subset \mathcal{R}(\mathbf{T})$ $\implies$ Schur complement $\mathbf{W} - \mathbf{V}\mathbf{T}^{-}\mathbf{U}$ is nonnegative. (Similar applies to Schur complement of relative to $\mathbf{W}^{-}$.)
  - **T14.8.5**. Partial converse of T14.8.4.

- Necessary and/or sufficient conditions for positive/nonnegative definiteness of symmetric matrices

- **T14.9.5**. Let $\mathbf{A}$ be symmetric. $\mathbf{A}$ is positive definite $\iff$ the determinants of all of its leading principal submatrices are positive.
- **T14.9.7**. Let $\mathbf{A}$ be symmetic. $\mathbf{A}$ is positive definite $\iff$ (1) is nonsingular (2) the determinants of all its principal submatrices are nonnegative.
- **L14.9.10**. Let $\mathbf{A}$ be symmetric of rank r with a non-singular $r \times r$ principal submatrix $\mathbf{A}_*$. $\mathbf{A}$ is nonnegative $\iff$ $\mathbf{A}_*$ is positive definite.
- **T14.9.1**. Let $\mathbf{A}$ be symmetric. $\mathbf{A}$ is nonnegative definite $\iff$ Every principal submatrix of $\mathbf{A}$ is nonnegative.

- Geomatrical considerations.

  - p253. For any symmetric positive $\mathbf{A}$, $\mathbf{x}'\mathbf{A}\mathbf{y}$ is an inner product for $\mathcal{R}^{n \times 1}$ as it satisfies all 4 properties (in p59).
  - p253. Any inner product for $\mathcal{R}^{n \times 1}$ is expressible as a bilinear form.

- Some connections with linear spaces and row/commmn spaces

  - **T14.11.1**. Let $\mathbf{W}$ be $m \times m$ symmetric nonnegative. Then, for any $m \times n$ $\mathbf{A}$ and $m \times p$ $\mathbf{B}$, the linear system $\mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{X} = \mathbf{A}'\mathbf{W}\mathbf{B}$ is consistent.
  - **L14.11.2**. For any $m \times n$ $\mathbf{A}$ and any $m \times m$ symmetric nonnegative $\mathbf{W}$, $\mathcal{C}(\mathbf{A}'\mathbf{W}\mathbf{A}) = \mathcal{C}(\mathbf{A}'\mathbf{W})$, $\mathcal{R}(\mathbf{A}'\mathbf{W}\mathbf{A}) = \mathcal{C}(\mathbf{W}\mathbf{A})$ and $\text{rank}(\mathbf{A}'\mathbf{W}\mathbf{A}) = \text{rank}(\mathbf{W}\mathbf{A})$.

- Projection, Inner product, Generalized Least Squares

  - **T14.2.8**. Let $\mathcal{U} \subset \mathcal{R}^n$, $\mathbf{X}$ be s.t. $\mathcal{C}(\mathbf{X}) = \mathcal{U}$, $\mathbf{W}$ be any spd. Then, $\exists$ A s.t. $\mathbf{A}\mathbf{y}$ is the projection of any $\mathbf{y} \in \mathcal{R}^n$ on $\mathcal{U}$. Moreover, $\mathbf{A} = \mathbf{P}_{\mathbf{X},\mathbf{W}}$.
  - **T14.12.11**. Properties of the projection matrix wrt spd $\mathbf{W}$ (more on p261)
    * $\mathbf{P}_{\mathbf{X},\mathbf{W}}\mathbf{X} = \mathbf{X}$ and also $\mathbf{W}\mathbf{P}_{\mathbf{X},\mathbf{W}}\mathbf{X} = \mathbf{W}\mathbf{X}$
    * $\mathbf{W}\mathbf{P}_{\mathbf{X},\mathbf{W}}$ and $\mathbf{P}_{\mathbf{X},\mathbf{W}}\mathbf{W}^{-1}$ are symmetric
    * $\mathbf{P}_{\mathbf{X},\mathbf{W}}$ and $(\mathbf{I} - \mathbf{P}_{\mathbf{X},\mathbf{W}})$ are idempotent
    * $\mathcal{C}(\mathbf{P}_{\mathbf{X},\mathbf{W}}) = \mathcal{C}(\mathbf{X})$ and $\mathcal{R}(\mathbf{P}_{\mathbf{X},\mathbf{W}}) = \mathcal{R}(\mathbf{X}'\mathbf{W})$
    * $\mathbf{P}'_{\mathbf{X},\mathbf{W}}\mathbf{W}\mathbf{P}_{\mathbf{X},\mathbf{W}} = \mathbf{W}\mathbf{P}_{\mathbf{X},\mathbf{W}}$
  - **T14.12.16**. A matrix $\mathbf{A}$ is a projection matrix wrt spd $\mathbf{W}$ $\iff$ $\mathbf{A}'\mathbf{W}\mathbf{A} = \mathbf{W}\mathbf{A}$.
  - **T14.12.18**. Any of the following are necessary and sufficient conditions for $\mathbf{P}_{\mathbf{X},\mathbf{V}} = \mathbf{P}_{\mathbf{X},\mathbf{W}}$
    1. $\mathbf{X}'\mathbf{V}\mathbf{P}_{\mathbf{X},\mathbf{W}} = \mathbf{X}'\mathbf{V}$
    2. $\mathbf{P}'_{\mathbf{X},\mathbf{W}}\mathbf{V}\mathbf{P}_{\mathbf{X},\mathbf{W}} = \mathbf{V}\mathbf{P}_{\mathbf{X},\mathbf{W}}$
    3. $\mathbf{V}\mathbf{P}_{\mathbf{X},\mathbf{W}}$ is symmetric
    4. $\exists$ matrix $\mathbf{Q}$ s.t. $\mathbf{V}\mathbf{X} = \mathbf{W}\mathbf{X}\mathbf{Q}$ or, equivalently, $\mathcal{C}(\mathbf{V}\mathbf{X}) \subset \mathcal{C}(\mathbf{W}\mathbf{X})$
  - **T14.12.23**. If $\mathbf{z}$ is a projection on orthogonal complement of $\mathbf{y}$ on $\mathcal{C}(\mathbf{X})^{\perp}_{\mathbf{W}}$, then $\mathbf{z} = \mathbf{y} - \mathbf{X}\mathbf{b}^*$ for any solution to $\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{W}\mathbf{y}$. In particular, $\mathbf{z} = \mathbf{y} - \mathbf{P}_{\mathbf{X},\mathbf{W}}\mathbf{y} = (\mathbf{I} - \mathbf{P}_{\mathbf{X},\mathbf{W}})\mathbf{y}$

Chapter 15

Matrix Differentiation

15.1   Notations

Differentiation of *scalar-valued* functions: Let $f(\mathbf{x})$ be a scalar-valued function of an m-dimensional vector $\mathbf{x}$.

- $\partial f(\mathbf{x})/\partial x_j$ or $D_j f(\mathbf{x})$ are the first order partial derivatives of f wrt x_j.
- $\partial f(\mathbf{x})/\mathbf{x}'$ or $\mathbf{D}f(\mathbf{x})$ are row vectors $\partial f(\mathbf{x})\mathbf{x}' = \mathbf{D}f(\mathbf{x}) = (\partial f(\mathbf{x})/\partial x_1, \partial f(\mathbf{x})/\partial x_2, \ldots, \partial f(\mathbf{x})/\partial x_m)$
- $\partial f(\mathbf{x})/\mathbf{x} = [\partial f(\mathbf{x})/\mathbf{x}']'$

Differentiation of *vector of functions*: Let $\mathbf{f}$ be a column vector with p functions $\mathbf{f} = (f_1, \ldots, f_p)'$.

- $D_j\mathbf{f}(\mathbf{c})$ is the column vector $D_j\mathbf{f}(\mathbf{c}) = [D_j f_1(\mathbf{c}), \ldots, D_j f_p(\mathbf{c})]'$.

Differentiation of matrix of functions: Let $\mathbf{F}$ be a $p \times q$ matrix of pq functions, $\mathbf{F} = \{f_{st}\}$.

- $\partial \mathbf{F}(\mathbf{x})/x_j$ is the $p \times q$ matrix whose stth element $\partial f_{st}(\mathbf{x})/x_j$.

15.2   Definitions

- *Neighborhood* of an m-dimensional vector $\mathbf{c}$ with the usual norm for $\mathcal{R}^{m \times 1}$ is $B_r(\mathbf{c}) = \{\mathbf{x} \in \mathcal{R}^{m \times 1} : ||\mathbf{x} - \mathbf{c}|| < r\}$.
- *Interior point.* If there is a neighborhood of $\mathbf{x}$ that is contained entirely in a set S, then $\mathbf{x}$ is an interior point of S.

Let $f(\mathbf{x})$ be a scalar-valued function of an m-dimensional vector $\mathbf{x}$.

- *Open set* is a set whose all points are interior points.
- *Continuity*: A function $f(\mathbf{x})$ is continuous at an interior point $\mathbf{c}$ if $\lim_{\mathbf{x} \to \mathbf{c}} f(\mathbf{x}) = f(\mathbf{c})$.
- *Continuously differentiable function* is a function whose all first-order derivatives exists and are continuous (p288).
- *Continuously differentiable matrix* is an $\mathbf{F}$ whose all pq elements (*i.e.* functions) are continuously differentiable.
- *Gradient vector* of f is $[\mathbf{D}f(\mathbf{x})]' = \partial f(\mathbf{x})/\mathbf{x}$
- *Hessian*, notated with $\mathbf{H}f$, is the $m \times m$ matrix whose ijth element is $\partial^2 f(\mathbf{x})/\partial x_i \partial x_j$

Let $\mathbf{f}$ be a column vector with p functions $\mathbf{f} = (f_1, \ldots, f_p)'$.

- *Jacobian matrix* is the $p \times m$ matrix $\mathbf{Df}(\mathbf{c}) = [D_1\mathbf{f}(c), \ldots, D_m\mathbf{f}(c)]$ (see above for $D_j\mathbf{f}(c)$).
- *Jacobian* is $\det[\mathbf{Df}(\mathbf{c})]$
- *Gradient vector* is the $m \times p$ matrix $[\mathbf{Df}(\mathbf{c})]'$

Some basic results:

- $\frac{\partial \mathbf{F}'}{x_j} = \left(\frac{\partial \mathbf{F}}{x_j}\right)'$

15.3 Results on matrix/vector differentiation

On linear and quadratic forms

- Let $\mathbf{a} = (a_1, \ldots, a_m)$ be constant (or fn of $\mathbf{x}$ that is invariant wrt x_j), and $\mathbf{A}$ be an $m \times m$ constant matrix (or matrix of functions invariant wrt x_j). Then:

  - $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial x_j} = a_j$ (see p294). The core idea is to see that $\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$

  - $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$ or $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}'$

  - $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_j} = \sum_{i=1}^{m} a_{ij} x_i + \sum_{k=1}^{m} a_{jk} x_k$ (see #295)

  - $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$ — unless $\mathbf{A}$ is symmetric, in which case $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$

  - $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_s x_j} = a_{sj} + a_{js}$ (see #295)

  - $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}\partial \mathbf{x}'} = (\mathbf{A} + \mathbf{A}')$ — unless $\mathbf{A}$ is symmetric, in which case $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}\partial \mathbf{x}'} = 2\mathbf{A}$

Differentiation of vector/matrix forms

- Let $\mathbf{x} = (x_1, \ldots, x_m)$ and $\mathbf{X}$ a matrix.

  - $\frac{\partial \mathbf{x}}{\partial x_i} = \mathbf{u}_j$
  - if $\mathbf{X}$ matrix of 'independent' variables x_{ij}:
    $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j'$
  - if $\mathbf{X}$ symmetric matrix:
    $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j' + \mathbf{u}_j \mathbf{u}_i'$

Differentiation of matrix of functions

- Let $\mathbf{F}, \mathbf{G}$ be matrices of functions

  - $\frac{\partial \mathbf{F}\mathbf{G}}{\partial x_j} = \mathbf{F}\frac{\partial \mathbf{G}}{\partial x_j} + \frac{\partial \mathbf{F}}{\partial x_j}\mathbf{G}$
  - $\frac{\partial \mathbf{F}\mathbf{G}\mathbf{H}}{\partial x_j} = \mathbf{F}\mathbf{G}\frac{\partial \mathbf{H}}{\partial x_j} + \mathbf{F}\frac{\partial \mathbf{G}}{\partial x_j}\mathbf{H} + \mathbf{F}\mathbf{G}\frac{\partial \mathbf{H}}{\partial x_j}$
  - if g is fn of $\mathbf{x}$: $\frac{\partial g\mathbf{F}}{\partial x_i} = \frac{\partial g}{\partial x_i}\mathbf{F} + g\frac{\partial \mathbf{F}}{\partial x_i}$

Trace and differentiation

- Let $\mathbf{F}$ be a matrix of functions of $\mathbf{x}$, $\mathbf{A}$ matrix of constants and $\mathbf{X}$ a matrix of variables $\mathbf{X} = \{x_{ij}\}$

  - $\frac{\partial \text{tr}(\mathbf{F})}{\partial x_j} = \text{tr}\left(\frac{\partial \mathbf{F}}{\partial x_j}\right)$
  - if $\mathbf{X}$ is matrix of 'independent' variables: $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j'$
  - if $\mathbf{X}$ symmetric matrix: $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j' + \mathbf{u}_j \mathbf{u}_i'$

Derivatives of determinants

- $\mathbf{X}, \mathbf{F}, \mathbf{A}$ defined just as above

  - $\frac{\partial \det(\mathbf{X})}{\partial x_{ij}} = \xi_{ij}$ where ξ_{ij} is the cofactor of x_{ij}
  - If $\mathbf{X}$ is matrix of 'independent' variables
    * $\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = [\text{adj}(\mathbf{X})]'$
    * $\frac{\partial \log \det \mathbf{X}}{\partial \mathbf{X}} = (\mathbf{X}^{-1})'$
  - If $\mathbf{X}$ is symmetric matrix
    * $\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = 2\text{adj}(\mathbf{X}) - \text{diag}(\xi_{11}, \ldots, \xi_{mm})$
    * $\frac{\partial \log \det \mathbf{X}}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \text{diag}(y_{11}, \ldots, y_{mm})$
  - $\frac{\partial \det(\mathbf{F})}{\partial x_j} = \text{tr}\left[\text{adj}(\mathbf{F})\frac{\partial \mathbf{F}}{\partial x_j}\right]' = $ (if $\mathbf{F}$ nonsingular) $= |\mathbf{F}|\text{tr}\left(\mathbf{F}^{-1}\frac{p\mathbf{F}}{\mathbf{x}_j}\right)$
  - $\frac{\partial \log \det \mathbf{F}}{\partial x_j} = \text{tr}\left(\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_j}\right)$

Derivatives of inverses and generalized inverses

- **F** defined as above

  - If **F** nonsingular: $\frac{\partial \mathbf{F}^{-1}}{\partial x_j} = \mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_j}\mathbf{F}^{-1}$

  - For any **F**, $\exists$ a generalized inverse **G** such that $\frac{\partial \mathbf{G}}{\partial x_j} = \mathbf{G}\frac{\partial \mathbf{F}}{\partial x_j}\mathbf{G}$ (T15.10.1)

  - For any **F** and any of its GIs **G**, $\mathbf{F}\frac{\partial \mathbf{G}}{\partial x_j}\mathbf{F} = \mathbf{FG}\frac{\partial \mathbf{F}}{\partial x_j}\mathbf{GF}$ (L15.10.5)

15.4 Notes

- **T15.7.1** (The Chain Rule). Let $\mathbf{h} = \{h_i\}$ be an $n \times 1$ vector of functions of variables $\mathbf{x} = (x_1, \ldots, x_m)$. Let g be a scalar-valued function of a vector of $\mathbf{y} = (y_1, \ldots, y_n)$. Define $f(\mathbf{x}) = g[\mathbf{h}(\mathbf{x})]$. Then, the jth partial derivative of f:

$$D_j f(\mathbf{c}) = \sum_{i=1}^{n} D_i g[\mathbf{h}(c)]D_j h_i(\mathbf{c}) = \mathbf{D}g[\mathbf{h}(\mathbf{c})]D_j\mathbf{h}(\mathbf{c})$$

Alternative notation:

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}\frac{\partial h_i}{\partial x_j} = \frac{\partial g}{\partial \mathbf{y}'}\frac{\partial \mathbf{h}}{\partial x_j}.$$

Differentiation wrt. vector:

$$\frac{\partial f}{\partial \mathbf{x}'} = \sum_{i=i}^{n} \frac{\partial g}{\partial y_i}\frac{\partial h_i}{\partial \mathbf{x}'} = \frac{\partial g}{\partial \mathbf{y}'}\frac{\partial \mathbf{h}}{\partial \mathbf{x}'}.$$

Differentiation of vector of functions $\mathbf{f}(\mathbf{x}) = \mathbf{g}[\mathbf{h}(\mathbf{x})]$ wrt. single variable:

$$\frac{\partial \mathbf{f}}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial \mathbf{g}}{\partial y_i}\frac{\partial h_i}{\partial x_j} = \frac{\partial \mathbf{g}}{\partial \mathbf{y}'}\frac{\partial \mathbf{h}}{\partial x_j}.$$

Differentiation of vector of functions wrt. vector:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}'} = \sum_{i=1}^{n} \frac{\partial \mathbf{g}}{\partial y_i}\frac{\partial h_i}{\partial \mathbf{x}'} = \frac{\partial \mathbf{g}}{\partial \mathbf{y}'}\frac{\partial \mathbf{h}}{\partial \mathbf{x}'}.$$

Differentiation of an $n \times r$ matrix $\mathbf{H} = \{h_{ij}\}$ of functions $f(\mathbf{x}) = g[\mathbf{H}(\mathbf{x})]$ wrt. single variable:

$$\frac{\partial \mathbf{f}}{\partial x_j} = \sum_{i=1}^{n}\sum_{s=1}^{r} \frac{\partial \mathbf{g}}{\partial y_i}\frac{\partial h_{is}}{\partial x_j} = \mathrm{tr}\left[\left(\frac{\partial g}{\partial \mathbf{Y}}\right)'\frac{\partial \mathbf{H}}{\partial x_j}\right]$$

15.5 Some multiple integrals

Let **W** be $n \times n$ symmetric positive definite matrix, **A** an $n \times n$ matrix and $\mathbf{k}, \mathbf{c}$ be $n \times 1$ vectors. From C14.3.13 (see above), we know that $\exists$ nonsingular **P** s.t. $\mathbf{W} = \mathbf{P}'\mathbf{P}$. Using the variable transformation $\mathbf{z} = \mathbf{P}(\mathbf{x} - \mathbf{c})$, the following integrals are evaluated as:

$$\int_{\mathcal{R}^n} \exp[-(1/2)(\mathbf{x} - \mathbf{c})'\mathbf{W}(\mathbf{x} - \mathbf{c})]d\mathbf{x} = (2\pi)^{n/2}|\mathbf{W}|^{-1/2}$$

$$\int_{\mathcal{R}^n} (\mathbf{k}'\mathbf{x})\exp[-(1/2)(\mathbf{x} - \mathbf{c})'\mathbf{W}(\mathbf{x} - \mathbf{c})]d\mathbf{x} = (\mathbf{k}'\mathbf{c})(2\pi)^{n/2}|\mathbf{W}|^{-1/2}$$

$$\int_{\mathcal{R}^n} (\mathbf{x} - \mathbf{c})'\mathbf{W}(\mathbf{x} - \mathbf{c})\exp[-(1/2)(\mathbf{x} - \mathbf{c})'\mathbf{W}(\mathbf{x} - \mathbf{c})]d\mathbf{x} = (2\pi)^{n/2}|\mathbf{W}|^{-1/2}\mathrm{tr}(\mathbf{AW}^{-1})$$

Chapter 16

Intersections and Sums of Subspaces

The concept of essentially disjoint subspaces allows us to identify conditions to more precisely specify the rank of partitioned matrices or of multiplication of matrices. It also allows us to describe column or row spaces more specifically. Essential disjointness is closely related to projection as well.

16.1 Definitions

- *Sum of subspaces.* The sum of two subspaces $\mathcal{U}$ and $\mathcal{V}$ is defined as $\mathcal{U} + \mathcal{V} = \{\mathbf{U} \in \mathcal{U}, \mathbf{V} \in \mathcal{V}\}$.

    - **T17.4.1.** $\dim(\mathcal{U} + \mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{V}) - \dim(\mathcal{U} \cap \mathcal{V})$

- *Essentially disjoint.* Two subspaces $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{W}$ are called essentially disjoint if $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$. Essential disjointness can be seen as a generalization of orthogonality; orthogonal subspaces are essentially disjoint (**L17.1.9**) but the converse is not in general true.

    - **L.17.1.3.** Let $\{\mathbf{U}_1, \ldots, \mathbf{U}_r\}$ be a linearly independent set in $\mathcal{U}$ and $\{\mathbf{V}_1, \ldots, \mathbf{V}_s\}$ be two sets of linearly independent matrices in $\mathcal{U}$ and $\mathcal{V}$. If $\mathcal{U}$ and $\mathcal{V}$ are essentially disjoint, then the combined set $\{\mathbf{U}_1, \ldots, \mathbf{U}_r, \mathbf{V}_1, \ldots, \mathbf{V}_s\}$ is also linearly independent.

- *Direct sum.* If $\mathcal{U}$ and $\mathcal{V}$ are essentially disjoint subspaces, then their sum is called a direct sum, denoted as $\mathcal{U} \oplus \mathcal{V}$

    - $\dim(\mathcal{U} \oplus \mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{V})$ (**C17.1.6**)

16.2 Results on Ranks of Partitioned Matrices and Column/Row Spaces

- **T17.2.4.** If $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are essentially disjoint, then

$$\mathrm{rank}(\mathbf{A}, \mathbf{B}) = \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B}).$$

  If they are not, then $\mathrm{rank}(\mathbf{A}, \mathbf{B}) < \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$. This is a sharpening of the results of **L4.5.7**.

- **C17.2.8.** For any appropriately sized matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}[(\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}]$ are essentially disjoint. Intuitively terms, this result says that $\mathcal{C}[(\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}]$ "hits" the part of the $\mathbf{B}$'s column space that cannot be hit by $\mathcal{C}(\mathbf{A})$. A direct consequence of this is that

$$\mathrm{rank}[\mathbf{A}, (\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}] = \mathrm{rank}(\mathbf{A}) + \mathrm{rank}[(\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}].$$

- **C17.2.9** For any appropriately sized matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathcal{C}(\mathbf{A}, \mathbf{B}) = \mathcal{C}[\mathbf{A}, (\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}] = \mathcal{C}(\mathbf{A}) \oplus \mathcal{C}[(\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}].$$

  .

- **L17.2.2** Suppose that $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are essentially disjoint and that $\mathcal{C}(\mathbf{X}) \subset \mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{X}) \subset \mathcal{C}(\mathbf{B})$, then $\mathbf{X} = \mathbf{0}$.

- **L17.2.3** If $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are essentially disjoint, then $\mathbf{A}\mathbf{K} = \mathbf{B}\mathbf{L} =)$ for any $\mathbf{K}, \mathbf{L}$.

- **L17.2.6** $\mathbf{A}$ is idempotent if and only if $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{I} - \mathbf{A})$ are essentially disjoint.

- **L17.2.7** For any $\mathbf{A}$, $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{I} - \mathbf{A}^-\mathbf{A})$ are essentially disjoint.

- **C17.2.12** If $\mathbf{A}\mathbf{A}^-\mathbf{B} = \mathbf{0}$ or $\mathbf{B}\mathbf{B}^-\mathbf{A} = \mathbf{0}$, then $\mathcal{C}(\mathbf{A}), \mathcal{C}(\mathbf{B})$ are essentially disjoint.

- **T17.2.17** For any $\mathbf{T}, \mathbf{U}, \mathbf{V}$

$$\text{rank} \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{0} \end{bmatrix} = \text{rank}(\mathbf{V}) + \text{rank}[\mathbf{U}, \mathbf{T}(\mathbf{I} - \mathbf{V}^- \mathbf{V})].$$

- **T17.2.19** For any $\mathbf{T}, \mathbf{U}, \mathbf{V}$

$$\text{rank} \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{0} \end{bmatrix} \leq \text{rank}(\mathbf{V}) + \text{rank}(\mathbf{U}) + \text{rank}(\mathbf{T})$$

with eq. holding iff $\mathcal{R}(\mathbf{T})$ and $\mathcal{R}(\mathbf{V})$ are essentially disjoint and $\mathcal{C}(\mathbf{T})$ and $\mathcal{C}(\mathbf{V})$ are also essentially disjoint.

16.3   Results on Linear Systems and on G. Inverses of Partitioned Matrices

- **T17.3.1**. Suppose that $\mathcal{C}(\mathbf{A_1})$ and $\mathcal{C}(\mathbf{A_2})$ are essentially disjoint, and that $\mathcal{C}(\mathbf{B_1}) \subset \mathcal{C}(\mathbf{A_1})$ and $\mathcal{C}(\mathbf{B_2}) \subset \mathcal{C}(\mathbf{A_2})$. Then, for the solution to the linear system $(\mathbf{A}_1, \mathbf{A}_2) \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \mathbf{B}_1 + \mathbf{B}_2$, $\mathbf{X}_1$ and $\mathbf{X}_2$ are solutions to the systems $\mathbf{A}_1 \mathbf{X}_1 = \mathbf{B}_1$ and $\mathbf{A}_2 \mathbf{X}_2 = \mathbf{B}_2$.

- **T17.3.3** Assume $\begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{pmatrix}$ is GI of $(\mathbf{A}, \mathbf{B})$. $\mathbf{G}_1$ is GI of $\mathbf{A}$ and $\mathbf{G}_2$ is GI of $\mathbf{B}$ iff $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are essentially disjoint.

- **T17.3.4** provides a more complicated set of conditions for the GI of partitioned matrix $\begin{pmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix}$

16.4   More results on ranks including on multiplication of matrices

- **T17.4.3**. $\text{rank}[(\mathbf{I} - \mathbf{A}\mathbf{A}^-)\mathbf{B}] = \text{rank}(\mathbf{B}) - \dim[\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B})]$

- **T17.4.2**. $\text{rank}(\mathbf{A}, \mathbf{B}) = \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - \dim[\mathcal{C}(\mathbf{A}) \cap \mathcal{C}(\mathbf{B})]$

- **T17.5.2**. $\text{rank}(\mathbf{A}\mathbf{B}) \geq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n$ with equality holding iff $(\mathbf{I} - \mathbf{B}\mathbf{B}^-)(\mathbf{I} - \mathbf{A}^-\mathbf{A}) = \mathbf{0}$

- **T17.5.4**. $\text{rank}(\mathbf{A}\mathbf{B}) = \text{rank}(\mathbf{B}) - \dim[\mathcal{C}(\mathbf{B}) \cap \mathcal{N}(\mathbf{A})] = \text{rank}(\mathbf{A}) - \dim[\mathcal{C}(\mathbf{B}') \cap \mathcal{N}(\mathbf{A}')]$

Chapter 17

Sums (and Differences) of matrices

This chapter serves for the following purposes:

- Give expression to simplify the inversion or computation of determinant for matrices of certain forms (*e.g.* $\mathbf{R} + \mathbf{STU}$)

- Give results to provide on ranks of sums of matrices

17.1   Results on Determinants

- **T18.1.1** (Determinants): $|\mathbf{R} + \mathbf{STU}| = |\mathbf{R}||\mathbf{T}||\mathbf{T}^{-1} + \mathbf{U} + \mathbf{R}^{-1}\mathbf{S}|$

- **C18.1.2**: $|\mathbf{I}_n + \mathbf{SU}| = |\mathbf{I}_m + \mathbf{US}|$

- **C18.1.3**: $|\mathbf{I}_n + \mathbf{su}'| = 1 + \mathbf{s}'\mathbf{u}$

- **T18.1.6**: For symmetric p.d. $\mathbf{A}$ and symmetric n.n.d. $\mathbf{B}$, we have $|\mathbf{A} + \mathbf{B}| \geq |\mathbf{A}|$ w/ equality holding iff $\mathbf{B} = \mathbf{0}$

17.2   Results on Inverses

- **T18.2.4**: For nonsingular $\mathbf{A}, \mathbf{B}$, it holds that $\mathbf{A}^{-1} - \mathbf{B}^{-1}$ is nonsingular iff $\mathbf{A} - \mathbf{B}$ is nonsingular, in which case: $(\mathbf{A}^{-1} - \mathbf{B}^{-1})^{-1} = \mathbf{A} + \mathbf{A}(\mathbf{B} - \mathbf{A})^{-1}\mathbf{A}$

- **T18.2.8** (Woodbury's formula):

$$\begin{aligned}(\mathbf{R} + \mathbf{STU})^{-1} &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{S}(\mathbf{T}^{-1} + \mathbf{UR}^{-1}\mathbf{S})^{-1}\mathbf{UR}^{-1} \\ &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{ST}(\mathbf{T} + \mathbf{TUR}^{-1}\mathbf{ST})^{-1}\mathbf{TUR}^{-1}\end{aligned}$$

- **T18.2.9** (Special case of Woodbury): $(\mathbf{I}_n + \mathbf{SU})^{-1} = \mathbf{I}_n - \mathbf{S}(\mathbf{I}_m + \mathbf{US})^{-1}\mathbf{U}$

- **T18.2.16** (Inversion with convergent finite series): $(\mathbf{I} - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots$

- **T18.2.19**. T18.2.16 is useful only of the series $\sum_k \mathbf{A}^k$ converges, which clearly doesn't happen for matrices in general. This theorem gives sufficient conditions for this to happen.
  Let $\mathbf{A}$ be an $n \times n$ matrix. If $\|\|\mathbf{A}\|\| < 1$ (where the norm is $\sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$), then $\lim_{k \to \infty} \mathbf{A}^k = 0$.

- **L18.2.20**. Similarly, this is another way to check convergence for a series of matrices. The series $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots$ converges to a matrix $\mathbf{A}$ iff $\|\mathbf{A}_k - \mathbf{A}\| \to 0$ (where norm is as above).

17.3   Results on Ranks

- **C18.5.2**: $\mathrm{rank}(\mathbf{R} + \mathbf{STU}) = \mathrm{rank}(\mathbf{R}) + \mathrm{rank}(\mathbf{T} + \mathbf{TUR}^{-1}\mathbf{ST}) - \mathrm{rank}(\mathbf{T})$

- **T18.5.7**: For any matrices $\mathbf{A}, \mathbf{B}$, $\mathrm{rank}(\mathbf{A} + \mathbf{B}) = \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$ iff $\mathcal{R}(\mathbf{A}), \mathcal{R}(\mathbf{B})$ are essentially disjoint and $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{B})$ are essentially disjoint.

17.4 Redo:

Chapter 4
- Theorem 4.4.10

Chapter 11
- Exercise 3

Chapter 13
- Exercise 12.a
- Re-prove T13.3.4

Towards a cheatsheet

Non-singularity

- Any positive definite matrix is nonsingular (L14.2.8)

-

Quizzes

- How to check if a column vector $\mathbf{b}$ is in the column space (or range) of a matrix $\mathbf{A}$?