Handout: Matrix Algebra Review

Suggested readings: Hansen (2020), Appendix A.

Introduction

Vectors and matrices will be really convenient pieces of notation that we will use throughout the semester. Not only do we use the notation and arithmetic operations, but we also use important properties of inverses, determinants, and quadratic forms. This handout provides a summary of the most important definitions and results, including proofs if the arguments will be useful in other settings. Hopefully, you have seen most (or all) of these results before. We will not have time to discuss them in a lot of detail and these notes are mainly for self-study. Every now and then, we will discuss some of the definitions and results in class.

1 Matrices

Definition 1. An $n \times m$ *matrix* is a rectangular array that consists of nm elements arranged in n rows and m columns. That is, if $\mathbf{A}$ is an $n \times m$ matrix, we write

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nm} \end{pmatrix}.$$

Therefore, for $1 \leq i \leq n$ and $1 \leq j \leq m$, a_{ij} denotes the element of the matrix located in the i^{th} row and j^{th} column. Later on, it can also be convenient to denote this element by $[\mathbf{A}]_{ij}$. We refer to $n \times m$ as the dimensions of the matrix $\mathbf{A}$. We will denote by $\mathbf{A}_{i\cdot}$ the i^{th} row of the matrix and $\mathbf{A}_{\cdot j}$ the j^{th} column of the matrix.

We now define some important special cases.

Definition 2. An $n \times m$ matrix is *square* if $n = m$.

Definition 3. A square matrix is *diagonal* if $a_{ij} = 0$ for all $i \neq j$.

Remark 1. The *main diagonal* of a matrix consists of the elements a_{ii}. The other elements are the *off-diagonal* elements. Therefore, a square matrix is diagonal if and only if all of its off-diagonal elements are 0.

Definition 4. A square matrix is *upper triangular* if it has zero entries below the main diagonal.

Definition 5. A square matrix is *lower triangular* if it has zero entries above the main diagonal.

Definition 6. A square matrix is *symmetric* if $a_{ij} = a_{ji}$ for all i and j.

Definition 7. The *transpose* of an $n \times m$ matrix $\mathbf{A}$, denoted by $\mathbf{A}'$, is the $m \times n$ matrix whose element located in the i^{th} row and j^{th} column is a_{ji}. That is, $[\mathbf{A}']_{ij} = [\mathbf{A}]_{ji}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$.

Remark 2. The square matrix $\mathbf{A}$ is symmetric if and only if $\mathbf{A} = \mathbf{A}'$.

Remark 3. Notice that $(\mathbf{A}')' = \mathbf{A}$.

Definition 8. A matrix with only one column or only one row is called a *vector*. An $n \times 1$ matrix is a *column vector* and a $1 \times m$ matrix is a *row vector*. A 1×1 matrix is a *scalar*. By convention, when we refer to a vector, we will be referring to a column vector. When v is a $n \times 1$ vector, we write $v \in \mathbb{R}^n$.

Definition 9. The *identity matrix* of size n, denoted $\mathbf{I}_{n\times n}$, is a square diagonal $n \times n$ matrix where the main diagonal entries are all ones.

2 Arithmetic Operations of Matrices

We can define the following algebraic operations for matrices.

Matrix addition: This operation works in the same way as for scalars, i.e. we add and subtract matrices element by element. In order to add two matrices we need them to have the same dimensions, in which case we say that the two matrices are *conformable for addition*. If $\mathbf{A}$ and $\mathbf{B}$ are conformable for addition, then $[\mathbf{A}+\mathbf{B}]_{ij} = [\mathbf{A}]_{ij} + [\mathbf{B}]_{ij} = a_{ij} + b_{ij}$. Recall that $[\mathbf{A}+\mathbf{B}]_{ij}$ denotes the element of $\mathbf{A}+\mathbf{B}$ located in the i^{th} row and j^{th} column. Subtraction of matrices is defined in the analogous way: $[\mathbf{A}-\mathbf{B}]_{ij} = [\mathbf{A}]_{ij} - [\mathbf{B}]_{ij}$. Matrix addition is commutative and associative, i.e.

$$
\begin{aligned}
\mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A} \\
(\mathbf{A}+\mathbf{B})+\mathbf{C} &= \mathbf{A}+(\mathbf{B}+\mathbf{C}).
\end{aligned}
$$

Scalar multiplication: This operation works in the same way as for scalars. If c is a scalar and $\mathbf{A}$ is a matrix, then a element (i,j) of $c\mathbf{A}$ is $[c\mathbf{A}]_{ij} = ca_{ij}$.

Matrix multiplication: The product, $\mathbf{AB}$, of two matrices $\mathbf{A}$ and $\mathbf{B}$ is only defined when the number of columns of the matrix $\mathbf{A}$ is equal to the number of rows of the matrix $\mathbf{B}$. In such a case, we say that $\mathbf{A}$ and $\mathbf{B}$ are *conformable for multiplication.* Notice that $\mathbf{A}$ and $\mathbf{B}$ being conformable for multiplication does not imply that $\mathbf{B}$ and $\mathbf{A}$ are conformable for multiplication because $\mathbf{BA}$ is only defined if the number of columns of the matrix $\mathbf{B}$ is equal to the number of rows of the matrix $\mathbf{A}$.

If $\mathbf{A}$ is an $n \times k$ matrix and $\mathbf{B}$ is a $k \times m$ matrix, then $\mathbf{AB}$ will be an $n \times m$ matrix where the element (i, j) is equal to $[\mathbf{AB}]_{ij} = \sum_{h=1}^{k} a_{ih} b_{hj}$. This operation is associative, in the sense that

$$(\mathbf{AB})\,\mathbf{C} = \mathbf{A}\,(\mathbf{BC})$$

but *not* commutative because in general $\mathbf{AB} \neq \mathbf{BA}$, even if both products are well defined. Notice that $\mathbf{BA}$ is well defined if $n = m$ in which case $\mathbf{AB}$ is an $n \times n$ matrix but $\mathbf{BA}$ is a $k \times k$ matrix.

Matrix addition and multiplication satisfy the distributive law, i.e.

$$\mathbf{A}\,(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

Matrix multiplication has an inverse operation associated with it, defined as follow.

Definition 10. The *inverse* of a square $n \times n$ matrix $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, is an $n \times n$ matrix such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_{n \times n}$$

Of course, not every matrix is invertible and, in particular, not every square matrix is invertible. We already knew this from the special case of scalars, where zero is not invertible, i.e. 0^{-1} is not defined. We will soon characterize necessary and sufficient conditions for the invertibility of a (square) matrix. We will also provide different properties of inverse matrices.

The following concepts are related to the product of matrices.

Definition 11. If $\mathbf{A}$ and $\mathbf{B}$ are column vectors, then the *inner product* between $\mathbf{A}$ and $\mathbf{B}$, is the (1×1) scalar $\mathbf{A}'\mathbf{B}$.

Definition 12. If $\mathbf{A}$ and $\mathbf{B}$ are column vectors, then the *outer product* between $\mathbf{A}$ and $\mathbf{B}$, is the $(n \times n)$ matrix $\mathbf{AB}'$.

Definition 13. A square matrix $\mathbf{A}$ is *idempotent* if $\mathbf{AA} = \mathbf{A}$. Notice that only square matrices can be idempotent.

We conclude by showing a few useful facts and more definitions.

Proposition 1. For any matrices $\mathbf{A}$ and $\mathbf{B}$, such that $\mathbf{A}$ and $\mathbf{B}$ are conformable for multiplication, $\mathbf{B}'$ and $\mathbf{A}'$ are also conformable for multiplication and

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

Proof. First, we show that $\mathbf{B}'$ and $\mathbf{A}'$ are conformable for multiplication . If $\mathbf{A}$ and $\mathbf{B}$ are conformable for multiplication, then $\mathbf{A}$ is $n \times k$ and $\mathbf{B}$ is $k \times m$. Therefore, $\mathbf{B}'$ is $m \times k$ and $\mathbf{A}'$ is $k \times n$, and so $\mathbf{B}'$ and $\mathbf{A}'$ are conformable for multiplication. Next, we show the equality. Consider the element $[\mathbf{AB}]_{ij}$ given by $\sum_{h=1}^{k} a_{ih} b_{hj}$. Therefore, we can obtain $[(\mathbf{AB})']_{ij}$ by interchanging i and j. That is, $[(\mathbf{AB})']_{ij} = \sum_{h=1}^{k} a_{jh} b_{hi}$. Now consider the element $[\mathbf{B}'\mathbf{A}']_{ij}$ given by $\sum_{h=1}^{k} [\mathbf{B}']_{ih} [\mathbf{A}']_{hj} = \sum_{h=1}^{k} b_{hi} a_{jh}$. Therefore, the elements $[\mathbf{AB}]'_{ij}$ and $[\mathbf{B}'\mathbf{A}']_{ij}$ coincide for all i and j, thus $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. □

We conclude this section with the definition of trace and its properties.

Definition 14. The *trace* of a square matrix $\mathbf{A}$, denoted $Tr\left(\mathbf{A}\right)$ is the sum of its diagonal elements

$$Tr\left(\mathbf{A}\right) = \sum_{i=1}^{n} a_{ii}$$

Proposition 2. The trace of a square matrix has the following properties:

1. For any scalar c, $tr\left(c\right) = c$.
2. For any scalar c and a matrix $\mathbf{A}$, $tr\left(c\mathbf{A}\right) = c tr\left(\mathbf{A}\right)$.
3. $tr\left(\mathbf{I}_n\right) = n$.
4. For any matrices $\mathbf{A}$ and $\mathbf{B}$, $tr\left(\mathbf{A} + \mathbf{B}\right) = tr\left(\mathbf{A}\right) + tr\left(\mathbf{B}\right)$.
5. Let $\mathbf{A}$ and $\mathbf{B}$ be matrices such that $\mathbf{AB}$ and $\mathbf{BA}$ be conformable. Then $tr\left(\mathbf{AB}\right) = tr\left(\mathbf{BA}\right)$.

Proof. All but the last property are very easy to check, so we only show the last one. Consider the element $[\mathbf{AB}]_{ij}$ given by $\sum_{h=1}^{k} a_{ih} b_{hj}$ and $tr\left(\mathbf{AB}\right) = \sum_{i=1}^{n} \sum_{h=1}^{k} a_{ih} b_{hi}$. If we repeat the calculations with the matrix $\mathbf{BA}$, we will obtain that $tr\left(\mathbf{BA}\right) = \sum_{h=1}^{k} \sum_{i=1}^{n} b_{hi} a_{ih}$. Therefore, $tr\left(\mathbf{AB}\right) = tr\left(\mathbf{BA}\right)$. □

3 Rank

The concept of rank of a matrix is fundamental. In order to define it, we need the concept of linear dependence and independence.

Definition 15. The k vectors $v_1, v_2, \dots, v_k \in \mathbb{R}^n$ are said to be *linearly dependent* if we can write one of them as a linear combination of the others, i.e. if there exist $j \in \mathbb{N}$, with $1 \leq j \leq k$, and $k-1$ coefficients $\alpha_1, \alpha_2, ..., \alpha_{j-1}, \alpha_{j+1}, ..., \alpha_k$ such that

$$v_j = \sum_{i \neq j} \alpha_i v_i$$

Remark 4. There are two alternative and equivalent ways to define linearly dependent vectors:

1. The k vectors $v_1, v_2, ..., v_k \in \mathbb{R}^n$ are linearly dependent if there exist k coefficients $\beta_1, ..., \beta_k$, not all of them zero, such that

$$\sum_{i=1}^{k} \beta_i v_i = \vec{0}$$

   where $\vec{0}$ denotes the $n \times 1$ vector of zeros.

2. We can reformulate the last statement in matrix notation. The k vectors $v_1, v_2, ..., v_k \in \mathbb{R}^n$ are linearly dependent if there exists a $k \times 1$ vector of coefficients $\beta \neq \overrightarrow{0}$ such that,

$$\begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \mathbf{V}\beta = \vec{0}$$

   where $\mathbf{V}$ is the $n \times k$ matrix whose j^{th} column is the vector v_j.

We can now define linear independence as follows.

Definition 16. The k vectors $v_1, v_2, ..., v_k \in \mathbb{R}^n$ are said to be *linearly independent* if they are not linearly dependent.

The following definition is equivalent and sometimes more useful.

Definition 17. Let $v_1, v_2, ..., v_k$ be k vectors in $\mathbb{R}^n$ and let $\mathbf{V}$ be the $n \times k$ matrix whose j^{th} column is the vector v_j. Then $v_1, v_2, ..., v_k$ are linearly independent if

$$\mathbf{V}\beta = \vec{0} \qquad \text{implies} \qquad \beta = \vec{0},$$

where β is a vector of size k.

This concept allows us to define the rank of a matrix.

Definition 18. The *rank* of a matrix is the maximum number of linearly independent rows of a matrix. The rank of a matrix $\mathbf{A}$ is denoted by $rank\left(\mathbf{A}\right)$.

The rank has the following properties.

Proposition 3. Consider an arbitrary $n \times k$ matrix $\mathbf{A}$.

1. The maximum number of linearly independent rows and the maximum number of linearly independent columns coincide, uniquely defining the rank.

2. $rank(\mathbf{A}) \leq \min\{k, n\}$.

3. The rank is invariant to interchanging rows or interchanging columns.

Proof. Skipped. □

Theorem 1. *Let* $\mathbf{A}$ *be* $n \times n$ *matrix. Then* $rank(\mathbf{A}) = n$ *if and only if* $\mathbf{A}$ *is invertible.*

Proof. Skipped. □

Finally, we define a full rank matrix.

Definition 19. For any arbitrary $n \times k$ matrix $\mathbf{A}$, we say that the matrix $\mathbf{A}$ has *full rank* if $rank(\mathbf{A}) = \min\{k, n\}$.

4 Determinant and Inversion

In this section we provide additional conditions for the existence of the inverse of a matrix inverse. We first define the determinant of a *square* matrix. The definition of determinant is obtained recursively: that is, we define the determinant of a 1×1 matrix (a scalar) and then, for any $n \geq 1$, we define the determinant of an $n \times n$ matrix based on the definition of a (related) $(n-1) \times (n-1)$ matrix. We will denote the determinant of a matrix $\mathbf{A}$ by $\det(\mathbf{A})$.

Definition 20. The *determinant of a scalar* is the scalar itself.

Next, we introduce the concepts of minors and cofactors.

Definition 21. The ij^{th} *minor* of a square $n \times n$ matrix $\mathbf{A}$ (for $n > 1$), denoted M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix that results from removing the i^{th} row and the j^{th} column from $\mathbf{A}$. That is

$$M_{ij} = \det \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{i-1,1} & \ddots & a_{i-1,j-1} & a_{i-1,j+1} & \ddots & a_{i-1,n} \\ a_{i+1,1} & \ddots & a_{i+1,j-1} & a_{i+1,j+1} & \ddots & a_{i+1,n} \\ \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix}$$

Definition 22. The ij^{th} *cofactor* of a square $n \times n$ matrix $\mathbf{A}$ (for $n > 1$), denoted C_{ij}, is the number given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

We are now ready to define the determinant of a square matrix.

Definition 23. The *determinant* of a square $n \times n$ matrix $\mathbf{A}$ (for $n > 1$) is the number

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij} C_{ij}$$

where $j \in \{1, 2, ..., n\}$, or equivalently,

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} C_{ij}$$

where $i \in \{1, 2, ..., n\}$.

As a lemma, we can compute the determinant of a 2×2 matrix.

Lemma 1. The determinant of a 2×2 matrix $\mathbf{A}$ is given by

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{21}a_{12}$$

Some useful properties of determinants are summarized in the following theorem.

Theorem 2. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix.*

1. *If* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by multiplying each entry of some given row (or column) of* $\mathbf{A}$ *by the nonzero constant* α*, then* $\det(\mathbf{B}) = \alpha \det(\mathbf{A})$.
2. *If* $\mathbf{A}$ *has a row (or column) of zeros, then* $\det(\mathbf{A}) = 0$.
3. *If* $\mathbf{A}$ *is lower triangular or upper triangular of dimension* n*, then the determinant of* $\mathbf{A}$ *is the product of the main diagonal terms, i.e.* $\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii}$. *In particular,* $\det(\mathbf{I}_n) = 1$.

Proof. Skipped. $\square$

It is possible to provide a closed form expression for the inverse of a matrix in terms of cofactors and determinants, but deriving this expression is quite complicated and not really that useful for us. What will be more important is to know certain properties of inverses.

Proposition 4. Let $\mathbf{A}$ be a square $n \times n$ matrix. The following statements are equivalent:

1. $\mathbf{A}$ has full (column or row) rank, i.e. *rank* $(\mathbf{A}) = n$;

2. The column (row) vectors of $\mathbf{A}$ are linearly independent;

3. $\mathbf{A}$ is invertible;

4. $\det(\mathbf{A}) \neq 0$.

Proof. Skipped. □

Proposition 5. Let $\mathbf{A}$ and $\mathbf{B}$ be square, invertible $n \times n$ matrices. Then the following properties hold:

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$;

2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$;

3. $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$;

4. The inverse of a lower (upper) triangular matrix is also lower (upper) triangular.

Proof. The first one follows from the definition of the inverse:

$$\begin{aligned}
(\mathbf{AB})^{-1}\mathbf{AB} &= \mathbf{I}_n \\
(\mathbf{AB})^{-1}\mathbf{A} &= \mathbf{B}^{-1} \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1}
\end{aligned}$$

The same idea applies to the second one,

$$\begin{aligned}
(\mathbf{A}')^{-1}\mathbf{A}' &= \mathbf{I} \\
\mathbf{A}(\mathbf{A}')^{-1\prime} &= \mathbf{I}' = \mathbf{I} \\
(\mathbf{A}')^{-1\prime} &= \mathbf{A}^{-1} \\
(\mathbf{A}')^{-1} &= \left(\mathbf{A}^{-1}\right)'
\end{aligned}$$

We will skip the other ones. □

Corollary 1. If $\mathbf{A}$ is an invertible, symmetric $n \times n$ matrix, then $\mathbf{A}^{-1}$ is also symmetric. This is because,

$$\left(\mathbf{A}^{-1}\right)' = (\mathbf{A}')^{-1} = \mathbf{A}^{-1}$$

5 Quadratic Forms

In this section, we will assume that $\mathbf{A}$ is a square, symmetric $n \times n$ matrix.

Definition 24. $\mathbf{A}$ is *positive semi-definite* if for all $x \in \mathbb{R}^n$

$$x'\mathbf{A}x \geq 0.$$

Notice that $x'\mathbf{A}x$ is a scalar. $\mathbf{A}$ is *negative semi-definite* if for all $x \in \mathbb{R}^n$

$$x'\mathbf{A}x \leq 0.$$

Definition 25. $\mathbf{A}$ is *positive definite* if for all non-zero $x \in \mathbb{R}^n$

$$x'\mathbf{A}x > 0$$

$\mathbf{A}$ is *negative definite* if for all non-zero $x \in \mathbb{R}^n$

$$x'\mathbf{A}x < 0$$

We get the following result.

Lemma 2. If $\mathbf{A}$ is positive (negative) semi-definite, then its diagonal elements are all non-negative (non-positive). If $\mathbf{A}$ is positive (negative) definite, then its diagonal elements are all positive (negative).

Proof. Just apply the definition with $x = (0, 0, ..., 0, 1, 0, ..., 0)$ and move the location of the one in the vector. $\square$

Below are other properties of quadratic forms that we will use throughout the semester.

Proposition 6. For an arbitrary $n \times m$ matrix $\mathbf{B}$, the matrix $\mathbf{B}'\mathbf{B}$ (as well as $\mathbf{B}\mathbf{B}'$) is positive semi-definite. If $\mathbf{B}$ is $n \times m$ with $m \leq n$ and $rank(\mathbf{B}) = m$, then $\mathbf{B}'\mathbf{B}$ is positive definite.

Proof. By definition, $\mathbf{B}'\mathbf{B}$ (a square $m \times m$ matrix) is positive semi-definite if for all $x \in \mathbb{R}^m$,

$$x'\mathbf{B}'\mathbf{B}x \geq 0$$

which is true because

$$x'\mathbf{B}'\mathbf{B}x = (\mathbf{B}x)'(\mathbf{B}x) = \sum_{i=1}^{n} \left([\mathbf{B}x]_i\right)^2 \geq 0$$

For the second part, notice that $(\mathbf{B}x)'(\mathbf{B}x) = 0$ if and only if $\mathbf{B}x = \overrightarrow{0}$. If $\mathbf{B}$ has full rank, then $\mathbf{B}x = \overrightarrow{0}$ if and only if $x = \overrightarrow{0}$. Therefore, the inequality is strict if and only if x is a nonzero vector. This is exactly the definition of positive definite. $\square$

Proposition 7. If a square $n \times n$ matrix is positive or negative definite, then it has full rank or, equivalently, it is invertible.

Proof. Suppose that the matrix $\mathbf{B}$ does not have full rank. Then, there is a $y \neq \overrightarrow{0}$ such that $\mathbf{B}y = \overrightarrow{0}$. If we premultiply the matrix by y', then we found a $y \neq \overrightarrow{0}$ such that $y'\mathbf{B}y = 0$, which means that the matrix $\mathbf{B}$ can be neither positive nor negative definite. □

Note that the converse is not true. For example consider the matrix,

$$\mathbf{B} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

This matrix has full rank and it is invertible (the determinant is -1). But it can be neither positive nor negative definite since it has both positive and negative elements in the main diagonal.

Proposition 8. Suppose that $\mathbf{A}$ is positive definite, then there exists a nonsingular matrix with the same dimensions, denoted $\mathbf{A}^{1/2}$, such that $\left(\mathbf{A}^{1/2}\right)' \mathbf{A}^{1/2} = \mathbf{A}$.

Proof. Skipped. □

Proposition 9. If $\mathbf{A}$ is positive definite and invertible, then $\mathbf{A}^{-1}$ is also positive definite.

Proof. Skipped. □

6 Partitioned Matrices

The concept of partitioned matrices and its inverse will be useful to derive components of Least Squares estimators.

Any $m \times n$ matrix $\mathbf{A}$ can be divided into different submatrices. This division of the original matrix is called a partition. For example, the $m \times n$ matrix $\mathbf{A}$ can be partitioned in the following 2×2 matrix of sub-matrices,

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

where $\mathbf{A}_{11}$ is an $r \times s$ matrix, $\mathbf{A}_{12}$ is a $r \times (n-s)$ matrix, $\mathbf{A}_{21}$ is a $(m-r) \times s$ matrix and $\mathbf{A}_{22}$ is a $(m-r) \times (n-s)$ matrix. Here $\mathbf{A}_{ij}$ dis different from a_{ij} or $[\mathbf{A}]_{ij}$.

We are particularly interested in an expression of the inverse of a partitioned (invertible) matrix in terms of the submatrices in its partition.

Proposition 10. Let the square and invertible $n \times n$ matrix $\mathbf{A}$ be divided into the following partition of sub-matrices,

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

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are also invertible. Then,

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{M}^{-1} & -\mathbf{M}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{M}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{M}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{pmatrix}$$

where $\mathbf{M} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$.

Proof. Skipped. □

7 Rules of Matrix Differentiation

The goal of this section is to review some results about differentiation of functions defined by matrices. Regarding the connection with classical econometrics, these results will be useful in computing Least Squares estimators.

Before stating the main results, we need a few definitions that will help us express the derivative of a matrix.

Definition 26. For a differentiable function $f : \mathbb{R}^k \to \mathbb{R}$, the derivative of f at $x \in \mathbb{R}^k$, called the *Gradient vector*, is defined as the $k \times 1$ column vector

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \cdots \\ \frac{\partial f(x)}{\partial x_k} \end{pmatrix}$$

Definition 27. For a curve $f : \mathbb{R} \to \mathbb{R}^n$, the derivative of f at $x \in \mathbb{R}$, called the *Tangent vector*, is defined as the $n \times 1$ column vector

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x} \\ \frac{\partial f_2(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{pmatrix}$$

Definition 28. For a differentiable function $f : \mathbb{R}^k \to \mathbb{R}^n$, the derivative of f at $x \in \mathbb{R}^k$, called the *Jacobian matrix*, is given by the $n \times k$ matrix

$$\frac{\partial f(x)}{\partial x'} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_k} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \cdots & \frac{\partial f_n(x)}{\partial x_k} \end{pmatrix}$$

Definition 29. The second derivative of a twice differentiable real valued function $f(x) : \mathbb{R}^k \to \mathbb{R}$, called *Hessian matrix*, denoted $D^2 f(x)$, is defined as the $k \times k$ matrix

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_k} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_k \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_k \partial x_k} \end{pmatrix}$$

By Young's Theorem, this matrix is symmetric.

We can now state and prove the following result.

Proposition 11. Let $\mathbf{A}$ be an $n \times k$ matrix. Then, for any column vector $x \in \mathbb{R}^k$

$$\frac{\partial (\mathbf{A}x)}{\partial x'} = \frac{\partial (x'\mathbf{A}')}{\partial x'} = \mathbf{A}$$

Moreover, if $\mathbf{A}$ is $k \times k$, then for any column vector $x \in \mathbb{R}^k$

$$\begin{aligned} \frac{\partial (x'\mathbf{A}x)}{\partial x} &= (\mathbf{A}'+\mathbf{A})\, x \\ D^2 (x'\mathbf{A}x) &= \mathbf{A}'+\mathbf{A} \end{aligned}$$

Proof. For the first statement, consider the function $f : \mathbb{R}^k \to \mathbb{R}^n$, defined by $f(x) = \mathbf{A}x$. This is really a column vector of n functions (one for each row). The i^{th} function is given by $f_i(x) = \mathbf{A}_{i\cdot} x : \mathbb{R}^k \to \mathbb{R}$. Differentiability is trivially satisfied by f being a linear mapping. The Jacobian matrix is given by

$$\frac{\partial \mathbf{A}x}{\partial x'} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nk} \end{pmatrix} = \mathbf{A}$$

We repeat the argument with $f(x) = x'\mathbf{A}'$. Again, this can be thought of as a row of n functions $f_i(x) = x' [\mathbf{A}']_{\cdot i} = \mathbf{A}_{i\cdot} x : \mathbb{R}^k \to \mathbb{R}$. Therefore, the Jacobian is given by

$$\frac{\partial (x'\mathbf{A}')}{\partial x'} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nk} \end{pmatrix} = \mathbf{A}$$

Finally, for the square $k \times k$ matrix $\mathbf{A}$ and any column vector $x \in \mathbb{R}^k$, consider the function $f : \mathbb{R}^k \to \mathbb{R}$ defined by $f(x) = x'\mathbf{A}x = \sum_{j=1}^k \sum_{i=1}^k x_i a_{ji} x_j$. In such a case, the gradient is a $k \times 1$ vector. The m^{th} element is given by,

$$\begin{aligned} \frac{\partial (x'\mathbf{A}x)}{\partial x_m} &= \sum_{i \neq m} x_i a_{mi} + \sum_{j \neq m} a_{jm} x_j + 2a_{mm} x_m = \sum_{i=1}^k x_i a_{mi} + \sum_{j=1}^k a_{jm} x_j \\ &= x' (\mathbf{A}_{m\cdot})' + x'\mathbf{A}_{\cdot m} = (\mathbf{A}_{m\cdot} + \mathbf{A}'_{\cdot m})\, x \end{aligned}$$

Therefore,

$$\frac{\partial \left(x'\mathbf{A}x\right)}{\partial x} \quad = \quad \left(\mathbf{A}' + \mathbf{A}\right) x$$

Now we move on to second derivatives. Consider the case $f\left(x\right) = x'\mathbf{A}x$. From our previous derivation, $\frac{\partial(x'\mathbf{A}x)}{\partial x_m} = \left(\mathbf{A}_{m\cdot} + \mathbf{A}'_{\cdot m}\right) x$. Thus, $\frac{\partial^2(x'\mathbf{A}x)}{\partial x_m \partial x_h} = a_{mh} + a_{hm}$. It follows that,

$$D^2\left(x'\mathbf{A}x\right) = \mathbf{A}' + \mathbf{A}$$

completing the proof. □

References

Hansen, B. (2020). *Econometrics.*