This material comes from Hansen Appendix A.

Linear Regression Notes 1: Review of Matrix Algebra

H A.1

A **vector** a is a $k \times 1$ list of numbers. We will follow the convention of (primarily) using column vectors. That is,

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

If $k = 1$, then a is a scalar. A **matrix** $\mathbf{A}$ is a $k \times r$ rectangular array of numbers which we will write as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kr} \end{bmatrix}$$

I will typically capitalize and use bold-font to indicate a matrix in the course notes and will underline it on the board, e.g., $\underline{A}$ (since it is hard to write in bold on the board).

The **transpose** of a matrix, which will denote by $\mathbf{A}'$ is obtained by flipping the matrix on its diagonal. That is,

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1r} & a_{2r} & \cdots & a_{kr} \end{bmatrix}$$

Notice that $\mathbf{A}'$ is an $r \times k$ matrix. For a $k \times 1$ vector a, its tranpose, a', is a $1 \times k$ vector. For a scalar a, $a = a'$.

A matrix is **square** if $k = r$. A square matrix is **symmetric** if $\mathbf{A} = \mathbf{A}'$. A square matrix is **diagonal** if the off-diagonal elements are all zero. The **identity matrix** is the diagonal matrix where all the elements on the diagonal are equal to 1. It is common to denote the $k \times k$ identity matrix by

$$\mathbf{I}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Matrix Addition

H A.3

If two matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ (here the notation just means that a_{ij} and b_{ij} are elements of each matrix) have the same dimension, then they can be added, and

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

Matrix addition is commutative, that is, $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. It is also associative: $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$.

Matrix Multiplication

H A.4

Let c denote a scalar, then (we define) $\mathbf{A}c = c\mathbf{A} = (a_{ij}c)$. If a and b are both $k \times 1$ vectors, then their **inner product** is

$$a'b = a_1b_1 + a_2b_2 + \cdots + a_kb_k = \sum_{j=1}^{k} a_jb_j$$

Further, notice that $a'b = b'a$. a and b are said to be **orthogonal** if $a'b = 0$.

If $\mathbf{A}$ is $k \times r$ and $\mathbf{B}$ is $r \times s$ (that is, the number of columns of $\mathbf{A}$ is the same as the number of rows of $\mathbf{B}$), then $\mathbf{A}$ and $\mathbf{B}$ are said to be **conformable** and the matrix produce $\mathbf{AB}$ is defined as

$$\mathbf{AB} = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_s \end{bmatrix} = \begin{bmatrix} a_1'b_1 & a_1'b_2 & \cdots & a_1'b_s \\ a_2'b_1 & a_2'b_2 & \cdots & a_2'b_s \\ \vdots & \vdots & \ddots & \vdots \\ a_k'b_1 & a_k'b_2 & \cdots & a_k'b_s \end{bmatrix}$$

where, for example, $a_1' = (a_{11}, a_{12}, \ldots, a_{1r})$ (which is the first row of $\mathbf{A}$) and $b_1 = (b_{11}, b_{21}, \ldots, b_{r1})'$ is the first column of $\mathbf{B}$. Notice that the product is a $k \times s$ matrix.

Matrix multiplication is not commutative, i.e., in general, $\mathbf{AB} \neq \mathbf{BA}$. But it is associative: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$. And it is distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

Trace

H A.5

The **trace** of $k \times k$ square matrix $\mathbf{A}$ is the sum of its diagonal elements:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$$

Here are some useful properties of trace (where $\mathbf{A}$ and $\mathbf{B}$ are square matrices and c is a scalar):

1. $\text{tr}(c\mathbf{A}) = c\text{tr}(\mathbf{A})$
2. $\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$
3. $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
4. $\text{tr}(\mathbf{I}_k) = k$

Another useful property is that if $\mathbf{A}$ is $k \times r$ and $\mathbf{B}$ is $r \times k$, then $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. Unlike the previous results, this one is not obvious, so let's provide a quick proof:

$$\begin{aligned}
\text{tr}(\mathbf{AB}) &= \text{tr} \begin{bmatrix} a_1'b_1 & a_1'b_2 & \cdots & a_1'b_k \\ a_2'b_1 & a_2'b_2 & \cdots & a_2'b_s \\ \vdots & \vdots & \ddots & \vdots \\ a_k'b_1 & a_k'b_2 & \cdots & a_k'b_k \end{bmatrix} \\
&= \sum_{i=1}^{k} a_i'b_i \\
&= \sum_{i=1}^{k} b_i'a_i \\
&= \text{tr}(\mathbf{BA})
\end{aligned}$$

Rank and Inverse

H A.6

The rank of a $k \times r$ matrix (with $r \leq k$)

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_r \end{bmatrix}$$

written $\text{rank}(\mathbf{A})$, is the number of linearly independent columns of $\mathbf{A}$. $\mathbf{A}$ is said to have **full rank** if $\text{rank}(\mathbf{A}) = r$. Linear independence means that there is no non-zero $k \times 1$ vector c such that $\mathbf{A}'c = 0$. For example,

$$\text{rank} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = 1, \quad \text{rank} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = 2$$

so that the second matrix has full rank but the first matrix does not (notice that the second column equals the first column times 2 so that they are not linearly independent; alternatively, you can notice that $\mathbf{A}'c = 0$ for $c = (2, -1)')$.

A square $k \times k$ matrix $\mathbf{A}$ is **nonsingular** if $\text{rank}(\mathbf{A}) = k$ (i.e., if it has full rank). If $\mathbf{A}$ is nonsingular, then it has an **inverse** $\mathbf{A}^{-1}$ that satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_k$$

For two non-singular matrices $\mathbf{A}$ and $\mathbf{C}$, another useful property is that

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

that is, for a nonsingular matrix, you can swap the order of transpose and inverse. Another useful property is that

$$(\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}$$

These properties are the ones that we'll use often though Appendix A.6 has several additional properties of nonsingular matrices that may be useful as a reference at some point.

Positive definite matrices

H A.10

A $k \times k$ symmetric matrix $\mathbf{A}$ is said to be **positive semi-definite** if $c'\mathbf{A}c \geq 0$ for any non-zero, $k \times 1$ vector c; this is often written $\mathbf{A} \geq 0$. $\mathbf{A}$ is said to be **positive definite** if $c'\mathbf{A}c > 0$ for any non-zero, $k \times 1$ vector c; this is often written $\mathbf{A} > 0$.

The textbook lists a number of properties of a positive definite matrix. One of these that we will use is that, if $\mathbf{A} > 0$, then $\mathbf{A}$ is nonsingular, $\mathbf{A}^{-1}$ exists, and $\mathbf{A}^{-1} > 0$.

Another is that, if $\mathbf{A}$ is positive definite, then we can find a square root matrix $\mathbf{A}^{1/2}$ such that $\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$ where $\mathbf{A}^{1/2}$ is itself positive definite and symmetric.

Idempotent Matrices

H A.11

A $k \times k$ square matrix $\mathbf{A}$ is **idempotent** if $\mathbf{AA} = \mathbf{A}$.

Matrix Calculus

H A.20

For this section, let $x = (x_1, x_2, \ldots, x_k)'$ denote a $k \times 1$ vector and $g(x) : \mathbb{R}^k \rightarrow \mathbb{R}$. Now, let's consider taking the partial derivatives of the function g with respect to each variable in x; in particular,

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

I will typically follow the convention of taking vector derivatives like the previous one "down" (as above), but it is also useful to have a notation for taking vector derivatives "across" as in

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

Sometimes, we will also take second derivatives, which are given by

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

Notice that this is a $k \times k$ matrix which is symmetric and arises from taking the partial derivatives "down" and then "across".

Here are some examples (we will consider the case where a is a $k \times 1$ vector and $\mathbf{A}$ is a $k \times k$ symmetric matrix):

- $\frac{\partial}{\partial x}(a'x) = \frac{\partial}{\partial x}(x'a) = a$
- $\frac{\partial}{\partial x'}(\mathbf{A}x)_{k\times 1} = \mathbf{A} \quad \text{and} \quad \frac{\partial}{\partial x}(x'\mathbf{A})_{1\times k} = \mathbf{A}$
- $\frac{\partial}{\partial x}(x'\mathbf{A}x) = 2\mathbf{A}x \quad \text{and} \quad \frac{\partial}{\partial x'}(x'\mathbf{A}x) = 2x'\mathbf{A}$
- $\frac{\partial^2}{\partial x \partial x'}(x'\mathbf{A}x) = 2\mathbf{A}$

In my view, a main takeaway from the above examples is that matrix calculus behaves very much like scalar calculus as long as you pay close attention to keeping the dimensions of the matrices straight (and also pay some attention to where to put transposes).

Vec Operator and Kronecker Product

H A.21

Write the $k \times r$ matrix $\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_r \end{bmatrix}$. Then, the **vec** of **A** is defined as

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

which is a $kr \times 1$ vector that stacks all the columns of **A** into one long column.

Next, write $\mathbf{A} = (a_{ij})$, then the **Kronecker product** of **A** and **B** is defined as (note that there are not restrictions on the dimensions of the matrices):

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1r}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2r}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}\mathbf{B} & a_{k2}\mathbf{B} & \cdots & a_{kr}\mathbf{B} \end{bmatrix}$$

If the dimension of **B** is $m \times n$, then the dimension of $\mathbf{A} \otimes \mathbf{B}$ is $km \times rn$. The book provides some additional properties of Kronecker products.

Vector norms

H A.22

A **norm** is a function $\rho : \mathbb{R}^k \to \mathbb{R}$ that satisfies the following properties:

1. $\rho(ca) = c\rho(a)$ for any scalar c and $a \in \mathbb{R}^k$
2. $\rho(a + b) \leq \rho(a) + \rho(b)$. This is called the triangle inequality.
3. If $\rho(a) = 0$, then $a = 0$.

The three most common norm functions are

- The Euclidean norm: $||a|| = (a'a)^{1/2}$
- The 1-norm: $||a||_1 = \sum_{i=1}^{k} |a_i|$
- The sup-norm: $||a||_\infty = \max\{|a_1|, \ldots, |a_k|\}$