

# Linear Algebra

## Linear Dependence

Consider a set of  $k$   $n$ -dimensional vectors  $\{X_1, X_2, \dots, X_k\}$ . These vectors are,

**Definition 0.1.** *linearly dependent* if there exists a set of scalars  $\{a_1, a_2, \dots, a_k\}$  such that

$$a_1X_1 + a_2X_2 + \dots + a_kX_k = 0$$

where at least one  $a_i \neq 0$ .

Alternatively, they are,

**Definition 0.2.** *linearly independent* if the only set of scalars  $\{a_1, a_2, \dots, a_k\}$  that satisfies the above condition is  $a_1, a_2, \dots, a_k = 0$ .

If we collect these  $k$  column-vectors in a matrix,  $X = [X_1 \ X_2 \ \dots \ X_k]$ , then the linear dependence condition can be written as,

$$a_1X_1 + a_2X_2 + \dots + a_kX_k = [X_1 \ X_2 \ \dots \ X_k] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = Xa = 0$$

Given any  $n \times k$  matrix  $X$ , its columns are,

**Definition 0.3.** *linearly dependent* if there exists a vector  $a \in \mathbb{R}^k$  such that  $a \neq 0$  and  $Xa = 0$ ;

or,

**Definition 0.4.** *linearly independent* if the only vector  $a \in \mathbb{R}^k$  such that  $Xa = 0$  is  $a = 0$ .

For any matrix there may be more than one vector  $a \in \mathbb{R}^k$  such that  $Xa = 0$ . Indeed, if both  $a_1, a_2 \in \mathbb{R}^k$

satisfy this condition and  $a_1 \neq a_2$  then you can show that any linear combination of  $\{a_1, a_2\}$  satisfies the

condition  $X(a_1b_1 + a_2b_2) = 0$  for  $b_1, b_2 \in \mathbb{R}$ . Thus, there exists an entire set of vectors which satisfy this condition. This set is referred to as the,

**Definition 0.5.** *null space* of  $X$ ,

$$\mathcal{N}(X) = \{a \in \mathbb{R}^k : Xa = 0\}$$

It should be evident from the definition that if the columns of  $X$  are linearly independent then  $\mathcal{N}(X) = \{0\}$ , a singleton. That is, it just includes the 0-vector.

## Vector spaces, bases, and spans

Here, we concern ourselves only with real vectors from  $\mathbb{R}^n$ .

**Definition 0.6.** A *vector space*, denoted  $\mathcal{V}$ , refers to a set of vectors which is closed under finite addition and scalar multiplication.

**Definition 0.7.** A set of  $k$  linearly independent vectors,  $\{X_1, X_2, \dots, X_k\}$ , forms a *basis* for vector space  $\mathcal{V}$  if  $\forall y \in \mathcal{V}$  there exists a set of  $k$  scalars such that,

$$y = X_1 b_1 + X_2 b_2 + \dots + X_k b_k$$

Based on these definitions, it is evident that for the Euclidean space,  $\mathbb{E}^n$ , any  $n$  linearly independent vectors from  $\mathbb{R}^n$  is a basis. For example, any point in  $\mathbb{E}^2$  can be defined as a multiple of,

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Consider again the  $n \times k$  matrix  $X$ , where  $k < n$ . Then we define the,

**Definition 0.8.** *column space* (or *span*) of  $X$ , denoted  $\mathcal{S}(X)$ , as the vector space generated by the  $k$  columns of  $X$ . Formally,

$$\mathcal{S}(X) = \{y \in \mathbb{R}^n : y = Xb \text{ for some } b \in \mathbb{R}^k\}$$

A property to note about the span or column space  $X$  is,

**Result:**  $\mathcal{S}(X) = \mathcal{S}(XX') :::$

where  $XX'$  is a  $n \times n$  matrix.

Finally, we can define the,

**Definition 0.9.** *orthogonal column space* (or *orthogonal span*) of  $X$  as,

$$\mathcal{S}^\perp(X) = \{y \in \mathbb{R}^k : y'x = 0 \quad \forall x \in \mathcal{S}(X)\}$$

## Rank

Consider a  $n \times k$  matrix  $X$ , the

**Definition 0.10.** *row rank* of  $X$  is the maximum number of linearly independent rows:

$$\text{rowrank}(X) \leq n$$

We say that matrix  $X$  has *full* row rank if  $\text{rowrank}(X) = n$ .

The,

**Definition 0.11.** *column rank* of  $X$  is the maximum number of linearly independent columns:

$$\text{colrank}(X) \leq k$$

We say that matrix  $X$  has *full* column rank if  $\text{colrank}(X) = k$ .

An important result is,

- **Result:** the rank of  $X$ :

$$r(X) = \text{rowrank}(X) = \text{colrank}(X) \Rightarrow r(X) \leq \min\{n, k\}$$

In addition, since the  $r(X)$  depends on the number of linearly independent columns, we can say that,

- **Result:** the dimension of  $\mathcal{S}(X)$ ,  $\dim(\mathcal{S}(X))$ , is given by the  $r(X)$ .

Here are a few additional results,

- **Result:**  $r(X) = r(X')$
- **Result:**  $r(XY) \leq \min\{r(X), r(Y)\}$
- **Result:**  $r(XY) = r(X)$  if  $Y$  is square and full rank
- **Result:**  $r(X + Y) \leq r(X) + r(Y)$

## Properties of square matrices

Consider the case of a square,  $n \times n$ , matrix  $A$ . We say that,

**Definition 0.12.**  $A$  is *singular* if the  $r(A) < n$ ,

or that,

**Definition 0.13.**  $A$  is *non-singular* if the  $r(A) = n$ .

The singularity of a square matrix is important as it determines the invertibility of a matrix, which typically relates the existence of a unique solution in systems of linear equations. Here are a few key results,

- **Result:** There exists a matrix  $B = A^{-1}$ , such that  $AB = I_n$  (where  $I_n$  is the identity matrix), if and only if  $A$  is non-singular.
- **Result:**  $A$  is non-singular if and only if the determinant of  $A$  is non-zero:  $\det(A) \neq 0$ .<sup>1</sup>
- **Result:** Likewise,  $A$  is singular if and only if  $\det(A) = 0$ .
- **Result:**  $AA^{-1} = A^{-1}A = I$
- **Result:**  $(A')^{-1} = (A^{-1})'$
- **Result:** If their respective inverses exist, then  $(AB)^{-1} = B^{-1}A^{-1}$ .
- **Result:**  $\det(AB) = \det(A)\det(B)$
- **Result:**  $\det(A^{-1}) = \det(A)^{-1}$

For any square matrix  $A$ ,

**Definition 0.14.** the *trace* of  $A$  is the sum of all diagonal elements:

$$tr(A) = \sum_{i=1}^n a_{ii}$$

Regarding the trace of a square matrix, here are a few important results:

- **Result:**  $tr(A + B) = tr(A) + tr(B)$
- **Result:**  $tr(\lambda A) = \lambda tr(A)$  where  $\lambda$  is a scalar
- **Result:**  $tr(A) = tr(A')$

---

<sup>1</sup>These notes do not cover how to calculate the determinant of a square matrix. You should be able to find a definition easily online.

- **Result:**  $\text{tr}(AB) = \text{tr}(BA)$  where  $AB$  and  $BA$  are both square, but need not be of the same order.
- **Result:**  $\|A\| = (\text{tr}(A'A))^{1/2}$

## Properties of symmetric matrices

A symmetric matrix has the property that  $A = A'$ . Therefore,  $A$  must be square.

Here are a few important results concerning symmetric matrices.

- **Result:**  $A^{-1}$  exists if  $\det(A) \neq 0$  and  $r(A) = n$
- **Result:**  $A$  is *diagonalizable*.<sup>2</sup>
- **Result:** The eigenvector decomposition of a square matrix gives you  $A = C\Lambda C^{-1}$  where  $\Lambda$  is a diagonal matrix of eigenvalues and  $C$  a matrix of the corresponding eigenvectors. The symmetry of  $A$  gives you that  $C^{-1} = C' \Rightarrow A = C\Lambda C'$  with  $C'C = CC' = I_n$ .<sup>3</sup>

A key definition concerning symmetric matrices is their positive definiteness:

**Definition 0.15.**  $A$  is *positive semi-definite* if for any  $x \in \mathbb{R}^n$ ,  $x'Ax \geq 0$ .

Given the eigenvector decomposition of a symmetric matrix, *positive semi-definiteness* implies  $\Lambda$  is *positive semi-definite*:  $\lambda_i \geq 0 \quad \forall i$ . Likewise,

**Definition 0.16.**  $A$  is *positive definite* if for any  $x \in \mathbb{R}^n$ ,  $x'Ax > 0$ .

Again, based on the eigenvector decomposition, *positive semi-definiteness* implies  $\Lambda$  is *positive definite*:  $\lambda_i > 0 \quad \forall i$ .

A few more results are:

- **Result:**  $\text{tr}(A) = \sum_{i=1}^n \lambda_i$
- **Result:**  $r(A) = r(\Lambda)$
- **Result:**  $\det(A) = \prod_{i=1}^n \lambda_i$

---

<sup>2</sup>A matrix is diagonalizable if it is *similar* to some other diagonal matrix. Matrices  $B$  and  $C$  are similar if  $C = PBP^{-1}$ . A square matrix which is not diagonalizable is *defective*. This property relates closely to eigenvector decomposition.

<sup>3</sup>Recall, an eigenvalue and eigenvector pair,  $(\lambda, c)$ , of matrix  $A$  satisfy:

$$Ac = \lambda c \Rightarrow (A - \lambda I_n)c = 0$$

This last result can be used to prove that any positive definite matrix is non-singular and therefore has an inverse.

Any full-rank, positive semi-definite, symmetric matrix  $B$  has the additional properties:

- **Result:**  $B = C\Lambda C'$  and  $B^{-1} = C\Lambda^{-1}C'$
- **Result:** We can define the square-root of  $B$  as  $B^{1/2} = C\Lambda^{1/2}C'$ . Similarly,  $B^{-1/2} = C\Lambda^{-1/2}C'$ .

## Properties of idempotent matrices

An idempotent matrix has the property that  $D = DD$ . Therefore,  $D$  must be square.

Here are a few important results concerning idempotent matrices.

- **Result:**  $D$  is positive definite
- **Result:**  $D$  is diagonalizable
- **Result:**  $(I_n - D)$  is also an idempotent matrix
- **Result:** With the exception of  $I_n$ , all idempotent matrices are singular.
- **Result:**  $r(D) = \text{tr}(D) = \sum_{i=1}^n \lambda_i$
- **Result:**  $\lambda_i \in \{0, 1\} \quad \forall i$

*Projection* matrices are idempotent, but need not be symmetric. However, for the purposes of this module we will deal exclusively with symmetric idempotent projection matrices.

## Vector Differentiation

Here we will look at the derivatives of scalar with respect to (W.r.t.) a vector. You can also define other derivatives, such as the derivative of a vector w.r.t. a vector and the derivative of a scalar with respect to a matrix. However, these are not needed for these notes.

### General case

Suppose  $f(x) \in R$  (i.e. a scalar) and  $x \in R^n$  (i.e. a  $n \times 1$  vector). Then we can define the partial derivative of  $f(x)$  w.r.t.  $x$  as,

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

### Linear: scalar case

A special case is when  $f(x)$  is linear in  $x$ ,

$$f(x) = a'x = \sum_{i=1}^n a_i x_i$$

for  $a \in R^n$ . The derivative of  $a'x$  with respect to the **vector**  $x$  can be defined as,

$$\begin{aligned} \frac{\partial a'x}{\partial x} &= \begin{bmatrix} \frac{\partial a'x}{\partial x_1} \\ \frac{\partial a'x}{\partial x_2} \\ \vdots \\ \frac{\partial a'x}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \\ &= a \end{aligned}$$

since the the partial derivate of  $a'x = \sum_{i=1}^n a_i x_i$  w.r.t.  $x_i$  is just the scalar  $a_i$ .

The derivative of a scalar w.r.t. to a vector yields a vector of partial derivatives.

Since  $a'x$  is a scalar, it is by definition symmetric:  $a'x = x'a$ . Thus,

$$\frac{\partial x'a}{\partial x} = \frac{\partial a'x}{\partial x} = a$$

**Linear: vector case**

Suppose  $f(x)$  is a linear transformation of  $x$ ,

$$f(x) = A'x$$

for any  $m \times n$  matrix  $A$ ,

$$A = \begin{bmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_m \end{bmatrix}$$

where  $a_i \in R^n \forall i = 1, \dots, m$  and,

$$Ax = \begin{bmatrix} a'_1 x \\ a'_2 x \\ \vdots \\ a'_m x \end{bmatrix}$$

Note,  $f(x) = Ax \in R^m$ , a  $m \times 1$  vector. We can then define,

$$\begin{aligned} \frac{\partial Ax}{\partial x'} &= \begin{bmatrix} \frac{\partial a'_1 x}{\partial x_1} & \frac{\partial a'_1 x}{\partial x_2} & \cdots & \frac{\partial a'_1 x}{\partial x_n} \\ \frac{\partial a'_2 x}{\partial x_1} & \frac{\partial a'_2 x}{\partial x_2} & \cdots & \frac{\partial a'_2 x}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a'_m x}{\partial x_1} & \frac{\partial a'_m x}{\partial x_2} & \cdots & \frac{\partial a'_m x}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\ &= A \end{aligned}$$

Since  $Ax$  is  $m \times 1$  column vector, we take the derivative w.r.t.  $x'$  a row vector and not the column vector  $x$ . This results in a matrix of partial derivatives.

**Quadratic form**

A second special case is where the function takes on the quadratic form,

$$f(x) = x'Ax = \sum_{i=1}^N \sum_{j=1}^n a_{ij}x_i x_j$$



for  $n \times n$  (square) matrix  $A$ . As in the first linear case,  $f(x)$  is scalar.

Define  $c = Ax$ , the  $x'Ax = x'c$ . From the linear case, we know that,

$$\frac{\partial x'c}{\partial x} = c$$

Similarly, if we define  $d = A'x$  then  $x'Ax = d'x$ . From the linear case, we know that,

$$\frac{\partial d'x}{\partial x} = d$$

We can define the total derivative as the sum of the partial derivatives w.r.t. to the first and second  $x$ . Combining these two results, we have that,

$$\frac{\partial x'Ax}{\partial x} = Ax + A'x$$

And **if**  $A$  is symmetric, this result simplifies to  $2Ax$ .