

# IFT 6760A - Lecture 1

## Linear Algebra Refresher

**Scribe(s):** Tianyu Li

**Instructor:** Guillaume Rabusseau

### 1 Summary

In the previous lecture we have introduced some applications of linear algebra in machine learning, including linear regression, principle component analysis, as well as some basic concepts of spectral learning, method of moments. We also extended matrices to tensors, i.e. higher dimensional (larger than two) arrays. In addition, we also introduced the definition of *rank* for tensors and some sketches of the tensor decomposition formats.

In this lecture we will go back to the basics, refresh our memories of basic linear algebra. We will begin by introducing some basic definitions so that we have a sense of *vector space*, *subspace* and their *dimensions*. In the second part of this lecture, we will introduce the four types of subspace of a matrix.

### 2 Basic Definitions

**Definition 1** (Vector Space). A vector space  $(V, +, \cdot)$  over a field  $\mathbb{F}$  (e.g.  $\mathbb{R}, \mathbb{C}$ ) is a set  $V$  with two operations:

- Addition: if  $\mathbf{u}, \mathbf{v} \in V$  then  $\mathbf{u} + \mathbf{v} \in V$
- Scalar multiplication: if  $\alpha \in \mathbb{F}, \mathbf{v} \in V$ , then  $\alpha\mathbf{v} \in V$

A vector space  $V$  satisfies the following properties: for  $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \forall \alpha, \beta \in \mathbb{F}$

- Commutativity:  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- Associativity:  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- Additive identity:  $\exists \mathbf{0} \in V, \mathbf{u} + \mathbf{0} = \mathbf{u}$
- Additive inverse:  $\exists -\mathbf{u} \in V, \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- Distributive properties:  $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
- Distributive properties:  $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$
- Multiplicative identity:  $1\mathbf{v} = \mathbf{v}$  (where 1 is the identity of  $\mathbb{F}$ )

**Definition 2** (Subspace). Given a vector space  $V$ , a subset of  $V$ ,  $W \subseteq V$  is a subspace of  $V$  iff:

- $\mathbf{0} \in W$
- $\forall \mathbf{u}, \mathbf{v} \in W: \mathbf{u} + \mathbf{v} \in W$
- $\forall \alpha \in \mathbb{F}, \forall \mathbf{v} \in W: \alpha\mathbf{v} \in W$

**Definition 3** (Sum, Direct Sum). If  $W_1, W_2$  are two vector spaces, denote sum of  $W_1, W_2$  by  $W_1 + W_2$  and it is defined by:

$$W_1 + W_2 = \{\mathbf{w}_1 + \mathbf{w}_2 \mid \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2\}$$

In addition, if  $W_1 \subseteq V, W_2 \subseteq V$ , then  $W_1 + W_2$  is a subspace of  $V$ . Moreover, if  $V = W_1 + W_2$  and  $W_1 \cap W_2 = \{\mathbf{0}\}$ , we call the sum of  $W_1$  and  $W_2$  as the direct sum and is denoted by  $W_1 \oplus W_2$ .

**Definition 4** (Linear Combination). A linear combination of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in V$  is defined by:

$$\mathbf{v} = \sum_{i=1}^k \alpha_i \mathbf{v}_i$$

where  $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{F}$

In fact, we can stack all the vector in a column-wise fashion, i.e.

$$\mathbf{V} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \\ | & | & \cdots & | \end{bmatrix}$$

Then the linear combination  $\mathbf{v} = \sum_{i=1}^k \alpha_i \mathbf{v}_i$  can be written in its matrix form:  $\mathbf{v} = \mathbf{V}\mathbf{a}$ , where  $\mathbf{a}^\top = [\alpha_1, \alpha_2, \dots, \alpha_k]$

**Definition 5** (Span). The span of a set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in V$  is defined as:

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \left\{ \sum_{i=1}^k \alpha_i \mathbf{v}_i \mid \alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{F} \right\}$$

Note from the above definition,  $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$  is in fact a vector space. Moreover, it is a subspace of  $V$ , i.e.  $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) \subseteq V$ . We can also write Definition 5 in its matrix form. Following the matrix format above, we have:

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \{\mathbf{V}\mathbf{a} \mid \mathbf{a} \in \mathbb{F}^k\}$$

**Definition 6** (Linearly Independent). A set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in V$  is linearly independent if and only if (iff):

$$\sum_{i=1}^k \alpha_i \mathbf{v}_i = \mathbf{0} \Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

That is:

$$\mathbf{V}\mathbf{a} = \mathbf{0} \Rightarrow \mathbf{a} = \mathbf{0}$$

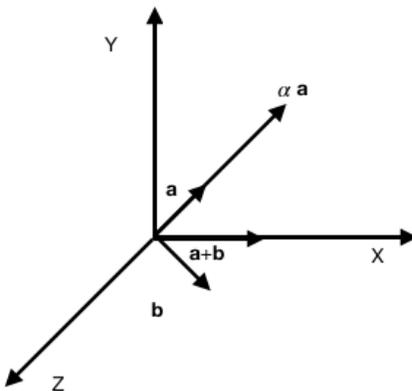
in the matrix form, where  $\mathbf{0} \in \mathbb{F}^k$  is an all zero vector.

**Definition 7** (Basis). Let  $W \subseteq V$  be a subspace of  $V$ ,  $w_1, w_2, \dots, w_k \in W$  form a basis of  $W$  iff:

- $\text{span}(w_1, w_2, \dots, w_k) = W$ .
- $w_1, w_2, \dots, w_k$  are linearly independent.

**Definition 8** (Dimension). The dimension of a subspace  $W \subseteq V$ , denoted by  $\dim(W)$ , is the number of vectors in any basis of  $W$ .

Take the most common Euclidean 3-D space as an example. Let us denote  $\mathbf{e}_1 = [1, 0, 0]^\top$ ,  $\mathbf{e}_2 = [0, 1, 0]^\top$ ,  $\mathbf{e}_3 = [0, 0, 1]^\top$  and denote the Euclidean 3-D space by  $D_3$ . One can easily check that  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  are a set of basis for  $D_3$ . In addition,  $\text{span}(\mathbf{e}_1, \mathbf{e}_2) = D_2$  is a subspace of  $D_3$ . In Figure 2, the plane (X, Y) is indeed the subspace  $D_2$ . Note that for any vector space, it is closed under the operation addition and scalar multiplication. Take  $D_2$  as an example, given  $\mathbf{a}, \mathbf{b} \in D_2, \alpha \in \mathbb{F}$ , then  $\mathbf{a} + \mathbf{b} \in D_2, \alpha\mathbf{a} \in D_2$ , as illustrated in Figure 2.

Figure 1: Illustration of  $D_2$  as a subspace of  $D_3$ 

### 3 Four Subspaces of a Matrix

From this section on, we will focus on the field of real numbers (i.e.  $\mathbb{R}$ ) instead of any arbitrary field. Since we will be introducing matrix in this section, it is good to clarify some notations beforehand.

As we did in the previous section, we will use bold lower case letters to denote vectors while bold capital letters to denote matrices. For  $\mathbf{v} \in \mathbb{R}^n$ ,  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , we use  $\mathbf{v}_i$  to denote the  $i$ th entry of the vector  $\mathbf{v}$  and  $\mathbf{M}_{i,j}$  to denote the  $(i, j)$ th entry of the matrix  $\mathbf{M}$ , where  $i \in [n]$ ,  $j \in [m]$  and  $[n] = \{1, 2, \dots, n\}$ . Similar to Python's notation, We denote the  $i$ th row of  $\mathbf{M}$  by  $\mathbf{M}_{i,:} \in \mathbb{R}^{1 \times n}$  and the  $j$ th column of  $\mathbf{M}$  by  $\mathbf{M}_{:,j} \in \mathbb{R}^{m \times 1}$  (by convention, a vector  $\mathbf{v} \in \mathbb{R}^m$  will always denote a column vector).

**Definition 9** (Matrix Product). For  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{n \times p}$ , the matrix product  $\mathbf{AB} \in \mathbb{R}^{m \times p}$  is defined by:

$$(\mathbf{AB})_{i,j} = \sum_{k=1}^n \mathbf{A}_{i,k} \mathbf{B}_{k,j} \quad \forall i \in [m], \forall j \in [p]$$

**Definition 10** (Transpose). For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , the transpose of the matrix  $\mathbf{M}$ ,  $\mathbf{M}^\top$  is defined by:

$$(\mathbf{M}^\top)_{i,j} = \mathbf{M}_{j,i} \quad \forall i \in [m], \forall j \in [n]$$

**Definition 11** (Inner Product). For  $\mathbf{a} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^n$ , the inner product of  $\mathbf{a}$  and  $\mathbf{b}$  is defined by:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b} = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i$$

**Definition 12** (Outer Product). For  $\mathbf{a} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^n$ , the outer product of  $\mathbf{a}$  and  $\mathbf{b}$  is defined by:

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \mathbf{b}^\top$$

That is, for  $i, j \in [n]$ :

$$(\mathbf{a} \otimes \mathbf{b})_{i,j} = \mathbf{a}_i \mathbf{b}_j$$

In fact, we can also define the matrix product in terms of inner product and outer product between vectors. We can

view each matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$  as either a stack of row vectors or a stack of column vectors, that is:

$$\mathbf{M} = \begin{bmatrix} | & | & & | \\ \mathbf{M}_{:,1} & \mathbf{M}_{:,2} & \cdots & \mathbf{M}_{:,n} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} -\mathbf{M}_{1,:} - \\ -\mathbf{M}_{2,:} - \\ \vdots \\ -\mathbf{M}_{m,:} - \end{bmatrix}$$

Then for the matrix product, we can either have the inner product view:

$$(\mathbf{AB})_{i,j} = \left( \begin{bmatrix} -\mathbf{A}_{1,:} - \\ -\mathbf{A}_{2,:} - \\ \vdots \\ -\mathbf{A}_{m,:} - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{B}_{:,1} & \mathbf{B}_{:,2} & \cdots & \mathbf{B}_{:,p} \\ | & | & & | \end{bmatrix} \right)_{i,j} = \langle \mathbf{A}_{i,:}^\top, \mathbf{B}_{:,j} \rangle$$

or we can have the outer product view:

$$\mathbf{AB} = \begin{bmatrix} | & | & & | \\ \mathbf{A}_{:,1} & \mathbf{A}_{:,2} & \cdots & \mathbf{A}_{:,n} \\ | & | & & | \end{bmatrix} \begin{bmatrix} -\mathbf{B}_{1,:} - \\ -\mathbf{B}_{2,:} - \\ \vdots \\ -\mathbf{B}_{n,:} - \end{bmatrix} = \sum_{i=1}^n \mathbf{A}_{:,i} \otimes \mathbf{B}_{i,:}^\top$$

Now we can finally define the four subspaces of a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .

**Definition 13 (Range).** The range (image, column space) of  $\mathbf{A}$  is the span of its columns, i.e.

$$\mathcal{R}(\mathbf{A}) = \text{span}(\mathbf{A}_{:,1}, \mathbf{A}_{:,2}, \dots, \mathbf{A}_{:,n}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$$

**Definition 14 (Nullspace).** The nullspace (kernel) of  $\mathbf{A}$  is defined by:

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$$

**Definition 15 (Row Space).** The row space of  $\mathbf{A}$  is the span of its rows, i.e.

$$\mathcal{R}(\mathbf{A}^\top) \subseteq \mathbb{R}^n$$

**Definition 16 (Left Nullspace).** The left nullspace of  $\mathbf{A}$  is defined by:

$$\mathcal{N}(\mathbf{A}^\top) \subseteq \mathbb{R}^m$$

Note that all these four subspaces are vector subspaces (they are closed under addition and scalar multiplication).

**Theorem 17.** The nullspace of a matrix is orthogonal to its row space:  $\mathcal{N}(\mathbf{A}) \perp \mathcal{R}(\mathbf{A}^\top)$ . That is, for all  $\mathbf{u} \in \mathcal{N}(\mathbf{A})$ ,  $\mathbf{v} \in \mathcal{R}(\mathbf{A}^\top)$  we have  $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

*Proof.* We want to show for all  $\mathbf{u} \in \mathcal{N}(\mathbf{A})$  and all  $\mathbf{v} \in \mathcal{R}(\mathbf{A}^\top)$ , we have  $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ . By Definition 15, there exists  $\mathbf{x} \in \mathbb{R}^m$  such that  $\mathbf{v} = \mathbf{A}^\top \mathbf{x}$ . Then we have  $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{A}^\top \mathbf{x} = (\mathbf{Au})^\top \mathbf{x}$ . Note by Definition 14, we have  $\mathbf{Au} = \mathbf{0}$ . Thus  $\langle \mathbf{u}, \mathbf{v} \rangle = (\mathbf{Au})^\top \mathbf{x} = 0$  and  $\mathcal{N}(\mathbf{A}) \perp \mathcal{R}(\mathbf{A}^\top)$ .  $\square$

**Definition 18 (Rank).** The rank of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the dimension of the range of  $\mathbf{A}$ :

$$\text{rank}(\mathbf{A}) = \dim(\mathcal{R}(\mathbf{A}))$$

**Definition 19** (Factorization). A factorization  $\mathbf{A} = \mathbf{BC}$ , where  $\mathbf{B} \in \mathbb{R}^{m \times r}$ ,  $\mathbf{C} \in \mathbb{R}^{r \times n}$  and  $r = \text{rank}(\mathbf{A})$ , is called a rank factorization. A graphical illustration of matrix factorization can be found in Figure 3.

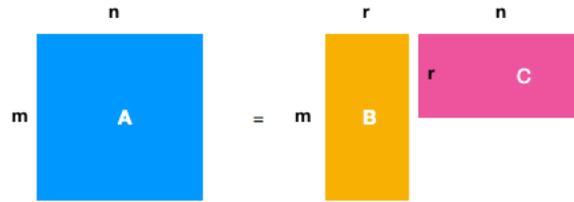


Figure 2: Illustration of matrix factorization

**Theorem 20.** For a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , all the followings are equivalent:

- (1)  $\text{rank}(\mathbf{A}) \leq r$
- (2)  $\dim(\mathcal{R}(\mathbf{A})) \leq r$
- (3)  $\exists \mathbf{B} \in \mathbb{R}^{m \times r}, \exists \mathbf{C} \in \mathbb{R}^{r \times n}$  such that:

$$\mathbf{A} = \mathbf{BC}$$

- (4)  $\exists \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r \in \mathbb{R}^m, \exists \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_r \in \mathbb{R}^n$  such that:

$$\mathbf{A} = \sum_{k=1}^r \mathbf{b}_k \mathbf{c}_k^\top$$

- (5)  $\dim(\mathcal{R}(\mathbf{A}^\top)) \leq r$

*Proof.* (1)  $\Leftrightarrow$  (2), by Definition 18.

(2)  $\Rightarrow$  (3): By Definition 8,  $\dim(\mathcal{R}(\mathbf{A})) \leq r$  indicates that there exists a set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in \mathbb{R}^m$  spanning  $\mathcal{R}(\mathbf{A})$ . Therefore, for  $i \in [n]$ ,  $\mathbf{A}_{:,i}$  can be written as a linear combination of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ . More precisely,  $\forall i \in [n]$ , there exists  $\gamma_{i,1}, \gamma_{i,2}, \dots, \gamma_{i,r}$  such that  $\mathbf{A}_{:,i} = \sum_{j=1}^r \gamma_{i,j} \mathbf{v}_j$ . Define the following matrices:

$$\mathbf{B} = \begin{bmatrix} | & | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \\ | & | & | & | \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \gamma_{1,1} & \gamma_{2,1} & \cdots & \gamma_{n,1} \\ \gamma_{1,2} & \gamma_{2,2} & \cdots & \gamma_{n,2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{1,r} & \gamma_{2,r} & \cdots & \gamma_{n,r} \end{bmatrix}$$

Then we have  $\mathbf{A} = \mathbf{BC}$ , where  $\mathbf{B} \in \mathbb{R}^{m \times r}$  and  $\mathbf{C} \in \mathbb{R}^{r \times n}$ .

(3)  $\Leftrightarrow$  (4): It is easy to check from the inner product, outer product views of the matrix product we mentioned earlier.

(4)  $\Rightarrow$  (5): If  $\mathbf{A} = \mathbf{BC}$ , where  $\mathbf{B} \in \mathbb{R}^{m \times r}$  and  $\mathbf{C} \in \mathbb{R}^{r \times n}$ , then it is easy to check by Definition 15, we have:

$$\text{span}(\mathbf{C}_{1,:}, \mathbf{C}_{2,:}, \dots, \mathbf{C}_{r,:}) = \mathcal{R}(\mathbf{A}^\top)$$

To see this, for  $i \in [m]$ , we have  $\mathbf{A}_{i,:} = \mathbf{B}_{i,:} \mathbf{C} = \sum_{k=1}^r \mathbf{B}_{i,k} \mathbf{C}_{k,:}$ . Thus  $\mathcal{R}(\mathbf{A}^\top) = \text{span}(\mathbf{C}_{1,:}, \mathbf{C}_{2,:}, \dots, \mathbf{C}_{r,:})$  and  $\dim(\mathcal{R}(\mathbf{A}^\top)) \leq r$ .

(5)  $\Rightarrow$  (1): The implication (5)  $\Rightarrow$  (3)  $\Rightarrow$  (2) follows from the exact same arguments reasoning by replacing  $\mathbf{A}$  with  $\mathbf{A}^\top$ . Since we have proved (1)  $\Leftrightarrow$  (2), we have (5)  $\Rightarrow$  (3)  $\Rightarrow$  (2)  $\Rightarrow$  (1) holds.  $\square$

**Theorem 21** (Rank Nullity). Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , then  $n = \text{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A}))$ .

*Proof.* Let  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r \in \mathbb{R}^m$  be a basis of  $\mathcal{R}(\mathbf{A})$ . Then, there exists  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r \in \mathbb{R}^n$  such that  $\forall i \in [r], \mathbf{u}_i = \mathbf{A}\mathbf{x}_i$ . Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$  be a basis of  $\mathcal{N}(\mathbf{A})$ . We want to show that  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  is a basis of  $\mathbb{R}^n$ .

First we show that  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  spans  $\mathbb{R}^n$ . For all  $\mathbf{y} \in \mathbb{R}^n$ , there exists  $\alpha_1, \alpha_2, \dots, \alpha_r$  such that  $\mathbf{A}\mathbf{y} = \sum_{i=1}^r \alpha_i \mathbf{u}_i = \sum_{i=1}^r \alpha_i \mathbf{A}\mathbf{x}_i$ , hence

$$\mathbf{A}(\mathbf{y} - \sum_{i=1}^r \alpha_i \mathbf{x}_i) = \mathbf{0}$$

Therefore, we have  $(\mathbf{y} - \sum_{i=1}^r \alpha_i \mathbf{x}_i) \in \mathcal{N}(\mathbf{A})$  and  $(\mathbf{y} - \sum_{i=1}^r \alpha_i \mathbf{x}_i) = \sum_{j=1}^k \beta_j \mathbf{v}_j$  for some scalars  $\beta_1, \dots, \beta_k$ . It follows that

$$\mathbf{y} = (\mathbf{y} - \sum_{i=1}^r \alpha_i \mathbf{x}_i) + \sum_{i=1}^r \alpha_i \mathbf{x}_i = \sum_{j=1}^k \beta_j \mathbf{v}_j + \sum_{i=1}^r \alpha_i \mathbf{x}_i$$

Therefore,  $\mathbb{R}^n = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ .

It remains to show that  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  are linearly independent. Suppose that  $\sum_{j=1}^k \beta_j \mathbf{v}_j + \sum_{i=1}^r \alpha_i \mathbf{x}_i = \mathbf{0}$ . This implies that  $\mathbf{A}(\sum_{j=1}^k \beta_j \mathbf{v}_j + \sum_{i=1}^r \alpha_i \mathbf{x}_i) = \mathbf{0}$  and thus that  $\mathbf{A}\sum_{i=1}^r \alpha_i \mathbf{x}_i = \sum_{i=1}^r \alpha_i \mathbf{u}_i = \mathbf{0}$  since  $\sum_{j=1}^k \beta_j \mathbf{v}_j \in \mathcal{N}(\mathbf{A})$ . Consequently  $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$  since the  $\mathbf{u}_i$  are linearly independent. To conclude, since  $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ , we have  $\mathbf{0} = \sum_{j=1}^k \beta_j \mathbf{v}_j + \sum_{i=1}^r \alpha_i \mathbf{x}_i = \sum_{j=1}^k \beta_j \mathbf{v}_j$ , hence  $\beta_1 = \beta_2 = \dots = \beta_k = 0$  since the  $\mathbf{v}_j$  are linearly independent.

Hence  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  form a basis of  $\mathbb{R}^n$ . Therefore,  $n = r + k = \text{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A}))$   $\square$

**Definition 22** (Linear Map). A map  $f : \mathbb{R}^n \mapsto \mathbb{R}^m$  is linear, if  $\forall \alpha, \beta \in \mathbb{R}, \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , we have:

$$f(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v})$$

**Remark 23** (Matrices and Linear Maps).

(1) Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , then the map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  defined by  $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$  for all  $\mathbf{x} \in \mathbb{R}^n$  is a linear map.

(2) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map and let  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  be the canonical basis of  $\mathbb{R}^n$ . Construct the matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$  such that:

$$\mathbf{M} = \begin{bmatrix} | & | & & | \\ f(\mathbf{e}_1) & f(\mathbf{e}_2) & \cdots & f(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}$$

then we have  $f(\mathbf{x}) = \mathbf{M}\mathbf{x}$  for all  $\mathbf{x} \in \mathbb{R}^n$ .

*Proof.* (1) is trivial to show.

(2) For all vectors  $\mathbf{x} \in \mathbb{R}^n$ , we have  $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$  (where  $x_i$  is the  $i$ th component of the vector  $\mathbf{x}$ ). Then  $\mathbf{M}\mathbf{x} = \sum_{i=1}^n x_i f(\mathbf{e}_i) = f(\sum_{i=1}^n x_i \mathbf{e}_i) = f(\mathbf{x})$  where we crucially used the linearity of  $f$  for the second equality.  $\square$

**Definition 24** (Kernel, Image). Given a linear map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the kernel of  $f$  is defined by:

$$\text{Ker}(f) = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = \mathbf{0}\}$$

The image of  $f$  is defined by:

$$\text{Im}(f) = \{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$$

**Remark 25.** Given a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , if  $f(\mathbf{x}) = \mathbf{M}\mathbf{x}$  for  $\mathbf{x} \in \mathbb{R}^n$ , then  $\text{Ker}(f) = \mathcal{N}(\mathbf{M})$  and  $\text{Im}(f) = \mathcal{R}(\mathbf{M})$