Linear Algebra

Chandra Gummaluru
chandra-gummaluru.github.io

Department of Electrical and Computer Engineering, University of Toronto

2022

Contents

1 Linear Spaces **3**
1.1 Linear Spaces and their Subspaces 3
1.2 Norms and Normed Linear Spaces 4
1.3 Inner Products and Inner Product Spaces 4

2 Linear Transformations **5**

3 Vectors and Linear Spaces of Vectors **7**

4 Matrices and Linear Spaces of Matrices **9**

1 Linear Spaces

1.1 Linear Spaces and their Subspaces

To define the notion of a linear space, we must first define the notion of a field. Loosely speaking, a field is any set endowed with a sense of addition and multiplication. A formal definition is given below.

Definition 1.1. (Field) A **field**, $\mathbb{F}$, is any set of elements endowed with the operations, $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ called addition, and $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$, called multiplication, such that the following properties are satisfied:

1. **commutativity**: $u + v = v + u$ and $u \cdot v = v \cdot u$ for all $u, v \in \mathbb{F}$.
2. **associativity**: $(u + v) + w = u + (v + w)$ and $(u \cdot v) \cdot w = u \cdot (v \cdot w)$ for all $u, v, w \in \mathbb{F}$.
3. **distributivity**: $u \cdot (v + w) = u \cdot v + u \cdot w$ for all $u, v, w \in \mathbb{F}$.
4. **identities**: there exists $0 \in \mathbb{F}$, called the additive identity, and $1 \in \mathbb{F}$ called the multiplicative identity, such that $v + 0 = 1 \cdot v = v$ for all $v \in \mathbb{F}$.
5. **inverses**: for every $v \in \mathbb{F}$, there exists $-v, v^{-1} \in \mathbb{F}$, called the additive and multiplicative inverses, such that $v + (-v) = 0$ and $v \cdot v^{-1} = 1$.

Linear spaces are defined over fields.

Definition 1.2. (Linear Space) A **linear space**, $\mathcal{V}$ over a field, $\mathbb{F}$, is a set of elements endowed with the operations $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$, called addition, and $\cdot : \mathcal{V} \times \mathbb{F} \to \mathcal{V}$, called scalar multiplication, such that the following properties are satisfied:

1. **commutativity of addition**: $u + v = v + u, \forall u, v \in \mathcal{V}$,
2. **associativity of addition**: $(u + v) + w = u + (v + w), \forall u, v, w \in \mathcal{V}$,
3. **distributivity**: $s(u + v) = su + sv$ and $u(s + t) = su + st, \forall u, v \in \mathcal{V}, s, t \in \mathcal{F}$,
4. **identities**: there exists $\mathbf{0} \in \mathcal{V}$ called the additive identity, and $1 \in \mathbb{F}$ called the multiplicative identity, such that $v + \mathbf{0} = 1v = v$ for all $v \in \mathcal{V}$,
5. **additive inverse**: $\forall v \in \mathcal{V}$ there exists $-v \in \mathcal{V}$, called the additive inverse, such that $v + (-v) = \mathbf{0}$.

While multiplication in $\mathbb{F}$ must be commutative and associative, its extension in $\mathcal{V}$ need not be. Similarly, while $\mathbb{F}$ must contain multiplicative inverses for all of its elements, $\mathcal{V}$ need not. Consequently, any field is a linear spaces over itself. However, not every linear space is a field.

Definition 1.3. (Linear Combination, Span, and Dimension) Let $\mathcal{X}$ be a linear space over a field, $\mathbb{F}$. A **linear combination** of $x_1, \ldots, x_m \in \mathcal{X}$ is of the form

$$\lambda_1 x_1 + \cdots + \lambda_m x_m, \lambda_i \in \mathbb{F}.$$

The **span** of $x_1, \ldots, x_m$, denoted $\text{span}(x_1, \ldots, x_m)$ is the set of all possible linear combinations of $x_1, \ldots, x_m$, i.e.,

$$\text{span}(x_1, \ldots, x_m) = \left\{ \sum_{i=1}^{m} \lambda_i x_i | \lambda_i \in \mathbb{F} \right\}.$$

The smallest m for which there exist a set of vectors, $x_1, \ldots, x_m \in \mathcal{X}$ such that $\text{span}(x_1, \ldots, x_m) = \mathcal{X}$ is called the **dimension** of $\mathcal{X}$, denoted $\dim(\mathcal{X})$.

Definition 1.4. (Linear Independence and Basis Sets) Let $\mathcal{X}$ be a linear space over a field, $\mathbb{F}$. We say $x^{(1)}, \ldots, x^{(m)} \in \mathcal{X}$ are **linearly independent** if $\lambda_1 x^{(1)} + \cdots + \lambda_m x^{(m)} = 0$ only when $\lambda_1 = \cdots = \lambda_m = 0$. A **basis** for $\mathcal{X}$ is any set of $\dim(\mathcal{X})$ linearly independent elements within $\mathcal{X}$ that also span $\mathcal{X}$.

We refer to $\lambda_1, \ldots, \lambda_m$ as the **coefficients** of x under the basis $\mathscr{X}$.

Generally speaking, linear spaces can be defined using any object. We will primarily focus on linear spaces of vectors, matrices, and functions. We will go into more detail about each of these objects in following sections. We will see that many of the linear spaces we work with also have additional structure which make them subspaces.

Definition 1.5. (Subspace) A linear space, $\mathcal{V}$, over a field $\mathbb{F}$, is a **subspace** if the following properties are satisfied:

1. **additive closure**: $u + v \in \mathcal{V}, \forall u, v \in \mathcal{V}$, and
2. **multiplicative closure**: $sv \in \mathcal{V}, \forall s \in \mathbb{F}, v \in \mathcal{V}$.

The beauty of a linear space that is also a subspace is that any properties enjoyed by its elements are preserved under the operations of the space. It turns out that both $\mathbb{R}$ and $\mathbb{C}$ are subspaces.

Definition 1.6. (Sum of Subspaces) Let $\mathcal{U}$, and $\mathcal{V}$ be subspaces of $\mathcal{X}$. We define their sum as

$$\mathcal{U} + \mathcal{V} := \{u + v, u \in \mathcal{U}, v \in \mathcal{V}\}.$$

If $\mathscr{U}$ is a basis for $\mathcal{U}$ and $\mathscr{V}$ is a basis for $\mathcal{V}$, then $\mathscr{U} \cup \mathscr{V}$ is a basis for $\mathcal{U} + \mathcal{V}$.

Definition 1.7. (Independent and Complement Subspaces) Let $\mathcal{U}$ and $\mathcal{V}$ be subspaces of $\mathcal{X}$. We say that $\mathcal{U}$ and $\mathcal{V}$ are **independent** if $\mathcal{U} \cap \mathcal{V} = \{\mathbf{0}\}$ and **complements** if $\mathcal{U} + \mathcal{V} = \mathcal{X}$.

It turns out that every subspace of a linear space has an independent complement.

1.2 Norms and Normed Linear Spaces

Definition 1.8. (Norm) Let $\mathcal{X}$ be a linear space over the field, $\mathbb{F}$. A function, $\|\cdot\| : \mathcal{X} \to \mathbb{F}$ is a **norm** iff the following properties are satisfied:

1. **sub-additivity**: $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{X}$.
2. **absolute homogeneity**: $\|sx\| = |s|\|x\|$ for all $x \in \mathcal{X}$ and $s \in \mathbb{F}$.
3. **non-negativity**: $\|x\| \geq 0$ for all $x \in \mathcal{X}$ and $\|x\| = 0 \Rightarrow x = \mathbf{0}$.

A linear space endowed with a norm is called a normed inner product space. We give a formal definition below.

Definition 1.9. (Normed Linear Space) A **normed linear space** is a linear space, $\mathcal{X}$ endowed with a norm, $\|\cdot\| : \mathcal{X} \to \mathbb{F}$.

1.3 Inner Products and Inner Product Spaces

Definition 1.10. (Inner-Product) Let $\mathcal{X}$ be a linear space over the field $\mathbb{F}$. An **inner product** is a function, $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \to \mathbb{F}$ such that:

1. **linearity**: $\langle sx, y \rangle = s\langle x, y \rangle$ and $\langle x + y, z \rangle = \langle x, y \rangle + \langle y, z \rangle$, for all $x, y, z \in \mathcal{X}$ and $s \in \mathbb{F}$.
2. **positivity**: $\langle x, x \rangle \geq 0$ if $x \neq \mathbf{0}$.
3. **conjugate symmetry**: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

These properties imply that $\langle x, x \rangle = 0$ if and only if $x = \mathbf{0}$.

Definition 1.11. (Orthogonality) Let $\mathcal{X}$ be an inner-product space. Two elements, $x, y \in \mathcal{X}$ are **orthogonal** iff $\langle x, y \rangle = \mathbf{0}$.

Theorem 1.1. (Mutual Orthogonality implies Linear Independence) Let $\mathcal{X}$ be an inner-product space over a field, $\mathbb{F}$. The elements $x^{(1)}, \ldots, x^{(n)} \in \mathcal{X}$ are linearly independent if and only if they are orthogonal

Proof. Assume $x^{(1)}, \ldots, x^{(n)}$ are mutually orthogonal, i.e., $\langle x^{(i)}, x^{(j)} \rangle = \mathbf{0}$ for all $i \neq j$, but that they are not linearly independent, i.e.,

$$\sum_{i=1}^{n} \lambda_i x^{(i)} = \mathbf{0}$$

and $\lambda_k \neq 0$ for some k. Taking the inner-product of both sides of the above equation with $x^{(j)}$, we have

$$\left\langle x^{(k)}, \sum_{i=1}^{n} \lambda_i x^{(i)} \right\rangle = \lambda_k \langle x^{(k)}, x^{(k)} \rangle = \langle x^{(k)}, \mathbf{0} \rangle = 0.$$

This is a contradiction since $\langle x^{(k)}, x^{(k)} \rangle \neq 0$. □

A linear space endowed with an inner product is called an inner product space. We give a formal definition below.

Definition 1.12. (Inner-Product Space) An **inner product space** is a linear space, $\mathcal{X}$, endowed with an inner-product $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \to \mathbb{F}$.

It turns out that any inner-product $\langle \cdot, \cdot \rangle$ induces a norm, $\| \cdot \|$ through the mapping $\|x\| = \langle x, x \rangle$, and thus every inner product space is also a normed space.

Theorem 1.2. (Orthogonal Complement) Let $\mathcal{V}$ be a subspace of an inner-product space $\mathcal{X}$. The **orthogonal complement** of $\mathcal{V}$, is

$$\mathcal{V}^\top := \{x \in \mathcal{X} : \langle x, v \rangle = 0, \forall v \in \mathcal{V}\}.$$

Proof. Clearly $\mathcal{V}$ and $\mathcal{V}^\top$ are independent. Indeed, $v \in (\mathcal{V}^\top \cap \mathcal{V}) \Leftrightarrow \langle v, v \rangle = 0 \Leftrightarrow v = 0$. □

2 Linear Transformations

In this section, we review linear transformations.

Definition 2.1. (Linear Transformation) Let $\mathcal{X}$ and $\mathcal{Y}$ be linear spaces over the field $\mathbb{F}$. A **linear transformation**, $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$, is a function satisfying

$$\mathbf{L}(x + \lambda y) = \mathbf{L}(x) + \lambda L(y),$$

for all $x, y \in \mathcal{X}$ and $\lambda \in \mathbb{F}$.

Definition 2.2. (Image, Kernel, and Rank) Let $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$ be a linear transformation. If $\mathcal{V}$ is a subspace of $\mathcal{X}$, then the image of $\mathcal{V}$ under $\mathbf{L}$ is

$$\mathbf{L}(\mathcal{V}) = \{y \in \mathcal{V} : (\exists x \in \mathcal{X}) y = \mathbf{L}(x)\}.$$

The image of $\mathcal{X}$ under $\mathbf{L}$, i.e., $\mathrm{L}(\mathcal{X})$, is called the **image** of $\mathbf{L}$, and is also denoted as $\operatorname{Img} L$. The **kernel** of $\mathbf{L}$, denoted $\operatorname{Ker} \mathbf{L}$, is the set of elements, $x \in \mathcal{X}$, such that $\mathbf{L}(x) = \mathbf{0}$, i.e.,

$$\operatorname{Ker} \mathbf{L} = \{x \in \mathcal{X} : \mathbf{L}(x) = \mathbf{0}\}.$$

It turns out that $\mathbf{L}(\mathcal{V})$ is a subspace of $\mathcal{Y}$ and $\operatorname{Ker} \mathbf{L}$ is a subspace of $\mathcal{X}$. The **rank** of $\mathbf{L}$ is the dimension of its image, i.e., $\operatorname{rank} \mathbf{L} = \dim(\operatorname{Img} \mathbf{L})$.

Definition 2.3. (Injective, Surjective, and Bijective Transformations) A linear transformation, $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$ is **injective** (or one-to-one) if

$$\mathbf{L}(x^{(1)}) = \mathbf{L}(x^{(2)}) \Leftrightarrow x^{(1)} = x^{(2)}, \forall x^{(1)}, x^{(2)} \in \mathcal{X},$$

surjective (or onto) if $\operatorname{Img} \mathbf{L} = \mathcal{Y}$, and **bijective** if it is both injective and surjective. If there exists some bijective transformation between $\mathcal{X}$ and $\mathcal{Y}$, then we say that $\mathcal{X}$ and $\mathcal{Y}$ are isomorphisms.

Theorem 2.1. (Injectiveness, Surjectiveness, Invertibility, and Rank) For any linear transformation, $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$, we have

$$\dim \operatorname{Img} \mathbf{L} + \dim \operatorname{Ker} \mathbf{L} = \dim \mathcal{X}.$$

Moreover, $\mathbf{L}$ is

- injective if and only if $\operatorname{Ker} \mathbf{L} = \{\mathbf{0}\}$, or equivalently, if $\dim \operatorname{Ker} \mathbf{L} = 0$,
- and surjective if and only if $\operatorname{rank} \mathbf{L} := \dim \operatorname{Img} \mathbf{L} = \dim \mathcal{Y}$.

It follows that if $\mathbf{L}$ is bijective, then $\dim \mathcal{X} = \dim \mathcal{Y}$. However, the converse is not necessarily true.

Proof. Omiitted. □

Suppose all elements in $\mathcal{X}$ are expressed using the basis $\mathscr{X} = \left\{x^{(1)}, \ldots, x^{(m)}\right\}$ and $x = \sum_{i=1}^{m} x_i x^{(i)}, x \in \mathcal{X}$. Our goal is to find an expression for L. To that end, consider applying L on x, i.e., $L(x)$. We have

$$\mathbf{L}(x) = \mathbf{L}\left(\sum_{i=1}^{m} x_i x^{(i)}\right) = \sum_{i=1}^{m} \mathbf{L}\left(x_i x^{(i)}\right) = \sum_{i=1}^{m} x_i \mathbf{L}\left(x^{(i)}\right).$$

In other words, the linear transformation of any element in $\mathcal{X}$ can be expressed as a linear combination of the linear transformations of the basis elements of $\mathcal{X}$. Now suppose that all elements in $\mathcal{Y}$ are expressed using the basis $\mathscr{Y} = \left\{y^{(1)}, \ldots, y^{(n)}\right\}$ and that $L\left(x^{(i)}\right) = \sum_{j=1}^{n} t_{j,i} y^{(i)}$. It follows that

$$\mathbf{L}(x) = \sum_{i=1}^{m} x_i \left(\sum_{j=1}^{n} t_{j,i} y^{(j)}\right) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} x_i t_{j,i}\right) y^{(j)}.$$

Observe that each coefficient of $\mathbf{L}(x)$ under $\mathscr{Y}$ is a linear combination of the coefficients of x under $\mathscr{X}$. We see that to define a linear transformation, $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$, using the bases $\mathscr{X}$ and $\mathscr{Y}$, we need to find the element in $\mathcal{Y}$ (expressed in $\mathscr{Y}$) that each basis element in $\mathcal{X}$ (expressed using $\mathscr{X}$) is mapped to.

Suppose we wish to represent the same transformation but using different bases. In particular, suppose we wish to use $\mathscr{A}$ as the basis for $\mathcal{X}$ and $\mathscr{B}$ as the basis for Y. If $\mathbf{L}_{\mathscr{X},\mathscr{Y}}$ represents the linear transformation in the original bases and $L_{\mathscr{A},\mathscr{B}}$ represents it in the new bases, we have

$$\mathbf{L}_{\mathscr{A},\mathscr{B}} = \mathbf{T}_{\mathscr{Y},\mathscr{B}} \circ \mathbf{L}_{\mathscr{X},\mathscr{Y}} \circ \mathbf{T}_{\mathscr{A},\mathscr{X}},$$

where $\mathbf{T}_{\mathscr{A},\mathscr{X}}$ is the change-of-basis transformation from $\mathscr{A}$ to $\mathscr{X}$ and $\mathbf{T}_{\mathscr{Y},\mathscr{B}}$ is the change-of-basis transformation from $\mathscr{Y}$ to $\mathscr{B}$. A change-of-basis transformation is always linear.

Definition 2.4. (Inverse Transform) Let $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$ be a bijective linear transformation. The **inverse** of $\mathbf{L}$, denoted $\mathbf{L}^{-1}$, is the unique linear transformation such that

$$\mathbf{L}^{-1} \circ \mathbf{L}(x) = x, \forall x \in \mathcal{X} \text{ and } L \circ \mathbf{L}(y) = y, \forall y \in \mathcal{Y}.$$

We note that the function $\mathbf{L}^{-1} : \mathcal{Y} \to \mathcal{X}$ is well-defined. Indeed, the surjectivity of $\mathbf{L}$ guarantees the existence of an x such that $\mathbf{L}(x) = y$, while its injectivity guarantees its uniqueness.

Definition 2.5. (Eigen-values and Eigen-elements) Let $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$ be some linear transformation, where $\mathcal{X}$ is a linear space over $\mathbb{F}$. An element, $x \in \mathcal{X}$ is an **eigen-element** of $\mathbf{L}$ if $\mathbf{L}(x) = \lambda x$ for some $\lambda \in \mathbb{F}$. We refer to λ as the associated **eigen-value**. In other words, the eigen-value/eigen-element pairs satisfy

$$\mathbf{L}(x) - \lambda x = \mathbf{0}.$$

For a fixed λ, we define

$$\text{Eig}(\mathbf{L}, \lambda) = \{x \in \mathcal{X} | \mathbf{L}(x) - \lambda x = \mathbf{0}\}$$

as the λ-**eigen-space** of $\mathbf{L}$. Since $\mathbf{L}$ is linear, $x \in \text{Eig}(\mathbf{L}, \lambda) \Leftrightarrow kx \in \text{Eig}(\mathbf{L}, \lambda), \forall k \in \mathbf{F}$. The **geometric multiplicity** of λ is the dimension of the λ-eigen-space, i.e., $\dim \text{Eig}(\mathbf{L}, \lambda)$.

3 Vectors and Linear Spaces of Vectors

We now consider vectors and linear spaces of vectors.

Definition 3.1. (Vector) Let n be a positive integer. An n-element column (resp. row) vector, v, over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$), is an n-tuple of elements in $\mathbb{F}$ stacked in a column (resp. row), i.e.,

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \left(\text{resp. } v = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}\right), v_i \in \mathbb{F}.$$

We write $v \in \mathbb{F}^n$. By convention, we assume (unless explicitly stated) that all vectors are column vectors. The set of all n-element (column) vectors over $\mathbb{F}$ is denoted $\mathbb{F}^n$.

Definition 3.2. (Vector Transpose) The **transpose** of an n-element column (resp. row) vector, v, denoted $v^\top$, is an n-element row (resp. column) vector containing the same elements, i.e.,

$$v = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \Leftrightarrow v^\top = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \text{ and } v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \Leftrightarrow v^\top = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}.$$

In order for $\mathbb{F}^n$ to be a linear space, we need to define the notions of addition between two vectors, and multiplication of a vector with a scalar. These definitions should be consistent with their scalar equivalents when $n = 1$.

Definition 3.3. (Vector Addition) Let $u, v \in \mathbb{F}^n$. We define their sum as

$$u + v := \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix},$$

where $u_i + v_i$ is defined implicitly by $\mathbb{F}$.

Definition 3.4. (Vector-Scalar Multiplication) Let $v \in \mathbb{F}^n$ and $s \in \mathbb{F}$. We define their product as

$$sv = \begin{bmatrix} sv_1 \\ \vdots \\ sv_n \end{bmatrix},$$

where sv_i is defined implicitly by $\mathbb{F}$.

With these operations, $\mathbb{F}^n$ becomes a linear space. This is expected since both operations are defined entirely in terms of their equivalents within the underlying field, which itself is a linear space. Moreover, $\mathbb{R}^n$ and $\mathbb{C}^n$ are subspaces of themselves.

All of these spaces have a dimensionality of n. So far, we have been representing vectors within them using the so-called **standard basis** for $\mathbb{F}^n$, i.e.,

$$\left\{\boldsymbol{e}^{(1)}, \dots, \boldsymbol{e}^{(n)}\right\},$$

where $\boldsymbol{e}^{(i)} \in \mathbb{F}^n$ is the vector with every element being the additive identity of $\mathbb{F}$, except the i^{th}, which is the multiplicative identity of $\mathbb{F}$. In other words, when we write $\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^\top$, we mean that $x = \sum_{i=1}^n x_i \boldsymbol{e}^{(i)}$. However, it is also possible to represent it using a different basis.

Definition 3.5. (Basis Representation) Let $\mathscr{X} = \{x^{(1)}, \dots, x^{(n)}\}$ be a (possibly non-standard) basis for the linear space $\mathcal{X}$ over the field $\mathbb{F}$. When we write $x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^\top_{\mathscr{X}}$ we mean

$$x = \sum_{i=1}^n x_i x^{(i)}.$$

Let us now define an inner-product for vectors.

Definition 3.6. (Vector Inner Product) We define the inner product of an n-element row vector, v, and n-element column vector w, as

$$vw = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \sum_{i=1}^{n} v_i w_i.$$

With this inner-product, $\mathbb{F}^n$ becomes an inner-product space. It is possible to define the vector inner-product in other ways. One use of defining the vector inner-product in the way we have is that it allows us to compactly express a linear equation, $\sum_{i=1}^{n} a_i x_i = b$, in vector form as $a^\top x = b$, where $a^\top = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}$ and $x^\top = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$.

4 Matrices and Linear Spaces of Matrices

We now consider matrices and linear spaces of matrices.

Definition 4.1. (Matrix) Let n and m be positive integers. An $n \times m$ dimensional matrix, A, over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$) is an array of n rows and m columns of elements in $\mathbb{F}$, i.e.,

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{n,1} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,m} \end{bmatrix}, a_{i,j} \in \mathbb{F}.$$

We write $A \in \mathbb{F}^{n,m}$. If $n = m$ we say that A is a **square matrix**. If A is a square matrix and $a_{i,j} = 0$ whenever $i \neq j$, we say that it is a **diagonal matrix** and often denote it concisely as $\text{diag}(a_{1,1}, \dots, a_{n,n})$.

In order for $\mathbb{F}^{n,m}$ to be a linear space, we need to define notions of addition between two matrices, and multiplication of a matrix with a scalar. These definitions should be consistent with their vector equivalents when $m = 1$.

Definition 4.2. (Matrix Addition) Let $A, B \in \mathbb{F}^{n,m}$. We define their sum as

$$A + B := \begin{bmatrix} a_{1,1} + b_{1,1} & \dots & a_{1,m} + b_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} + b_{n,1} & \dots & a_{n,m} + b_{n,m} \end{bmatrix}.$$

It is easy to see that if $m = 1$, the operation reduces to that of Def. **??**.

Definition 4.3. (Matrix-Scalar Multiplication) Let $A \in \mathbb{F}^{n,m}$ and $s \in \mathbb{F}$. We define their product as

$$sA := \begin{bmatrix} sa_{1,1} & \dots & sa_{1,m} \\ \vdots & \ddots & \vdots \\ sa_{n,1} & \dots & sa_{n,m} \end{bmatrix}.$$

It is easy to see that if $m = 1$, the operation reduces to that of Def. **??**.

With these operations, $\mathbb{F}^{n,m}$ becomes a linear space. This is expected since both operations are defined entirely in terms of their equivalents within the underlying field, which itself is a linear space. Moreover, $\mathbb{R}^{n,m}$ and $\mathbb{C}^{n,m}$ are subspaces of themselves.

All of these spaces have a dimensionality of mn. So far, we have been representing matrices using the so-called **standard basis** for $\mathbb{F}^{n,m}$, i.e.,

$$\left\{ \mathbf{e}^{(1,1)}, \dots, \mathbf{e}^{(n,1)}, \mathbf{e}^{(1,2)}, \dots, \mathbf{e}^{(m-1,n)}, \mathbf{e}^{(m,1)} \dots, \mathbf{e}^{(m,n)} \right\},$$

where $\mathbf{e}^{(i,j)} \in \mathbb{F}^{n,m}$ is the matrix with every element being the additive identity of $\mathbb{F}$, except for the element in the i^{th} row and j^{th} column which is the multiplicative identity of $\mathbb{F}$. In other words, when we write

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{m,1} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{m,n} \end{bmatrix},$$

we mean $A = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j} \mathbf{e}^{(i,j)}$. While it is also possible to represent it using a different basis, we will always use the standard basis.

Definition 4.4. (Columns and Rows of Matrices) Let $A \in \mathbb{F}^{n,m}$. The i^{th} column of A is an n-element column vector,

$$\begin{bmatrix} a_{1,i} \\ \vdots \\ a_{n,i} \end{bmatrix},$$

and the j^{th} row of A is an m-element row vector,

$$\begin{bmatrix} a_{j,1} & \dots & a_{j,m} \end{bmatrix}.$$

Using this interpretation, we can naturally define the product of two matrices in terms of the product of two vectors.

Definition 4.5. (Matrix-Matrix Multiplication) Let $A \in \mathbb{F}^{n,p}, B \in \mathbb{F}^{p,m}$. We define their product, $AB \in \mathbb{F}^{n,m}$, element-wise so that

$$(AB)_{i,j} = \sum_{k=1}^{m} a_{i,k} b_{k,j}, i = 1, \ldots, n, j = 1, \ldots, m.$$

In other words, the i^{th} row and j^{th} column of AB is the vector-vector product of the the i^{th} row of A and the j^{th} column of B.

Def. **??** holds still holds if A is a row vector, i.e., $n = 1$, or B is a column vector, i.e., $p = 1$, and consistent with Def. **??**. Furthermore, it allows us to compactly express a system of linear equations,

$$\begin{aligned} \sum_{i=1}^{m} a_{1,i} x_i &= b_1 \\ &\vdots \\ \sum_{i=1}^{m} a_{n,i} x_i &= b_n \end{aligned}$$

as $Ax = b$, where

$$A = \begin{bmatrix} a_{1,1} & \ldots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \ldots & a_{n,m} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \text{ and } b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Moreover, it lets us represent any linear transformation using a matrix. Consider $\mathbf{L} : \mathcal{X} \to \mathcal{Y}$ using the bases $\mathscr{X}$ and $\mathscr{Y}$. Recall that if $x = \sum_{i=1}^{m} x_i x^{(i)}$, then the j^{th} coefficient of $\mathbf{L}(x)$ is given by $\sum_{i=1}^{m} x_i t_{j,i}$, where $t_{j,i}$ satisfies $\mathbf{L}(x^{(i)}) = \sum_{j=1}^{n} t_{j,i} y^{(i)}$. In matrix form,

$$\mathbf{L}(x) = \underbrace{\begin{bmatrix} t_{1,1} & \ldots & t_{1,m} \\ \vdots & \ddots & \vdots \\ t_{n,1} & \ldots & t_{n,m} \end{bmatrix}}_{L_{\mathscr{X},\mathscr{Y}}} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

In other words, $L_{\mathscr{X},\mathscr{Y}}$ encodes the transformation $\mathbf{L}$ under $\mathscr{X}$ and $\mathscr{Y}$. We see that the i^{th} column of the $T_{\mathscr{X},\mathscr{Y}}$ is the representation of $x^{(i)}$ in $\mathscr{Y}$. However, this representation is not necessarily unique. Indeed, it depends on the choice of basis representing $\mathcal{X}$. If instead we sought the representation of $\mathbf{L}$ under $\mathcal{A}$ and $\mathcal{B}$, we can employ the change-of-bases transformations, i.e.,

$$L_{\mathscr{A},\mathscr{B}} = T_{\mathscr{Y},\mathscr{B}} L_{\mathscr{X},\mathscr{Y}} T_{\mathscr{A},\mathscr{X}}.$$

If $\mathbf{L}$ is an isomorphism, and $L_{\mathscr{X}}$ is its matrix representation under $\mathscr{X}$, and we seek its representation under $\mathscr{A}$, i.e., $L_{\mathscr{A}}$, we have

$$L_{\mathscr{A}} = T_{\mathscr{A},\mathscr{X}} L_{\mathscr{X},\mathscr{X}} T_{\mathscr{A},\mathscr{X}}.$$

Definition 4.6. (Similar Matrices) A square matrix, A, is **similar** to a square matrix X, denoted $A \equiv X$, if there exists a matrix T, such that $TA = XT$.

As we just saw, similar matrices represent the same transformation under different bases.

Definition 4.7. (Diagonalizable Matrix) A square matrix, A, is **diagonalizable** if it is similar to a diagonal matrix, i.e., there exists a diagonal matrix Δ, such that $TA = \Delta T$ for some matrix T.

Diagonal matrices enjoy a wide range of useful properties. Thus, it is often useful to determine when a matrix is diagonalizable and how to actually diagonalize it. To determine this, we will need to define a few additional concepts.

Definition 4.8. (Column Space, Row Space, and Rank) The **column space** of a matrix, A, denoted $\operatorname{col} A$, is the span of its columns, i.e., if

$$A = \begin{bmatrix} | & & | \\ a_1 & \vdots & a_m \\ | & & | \end{bmatrix}$$

then $\operatorname{col} A = \operatorname{span}\{a_1, \ldots, a_m\}$. The **row space** of A, denoted $\operatorname{row} A$, is the column space of $A^\top$.

The dimension of the column (resp. row) space, i.e., $\dim(\operatorname{col} A)$ (resp. $\dim(\operatorname{row} A)$) is called the **column** (resp. **row**) **rank** of A. It turns out that the column rank and row rank are always equal, i.e., $\dim(\operatorname{col} A) = \dim(\operatorname{row} A)$. Thus, we simply refer to either (or both) as the **rank** of A, denoted $\operatorname{rank} A$. An $n \times m$ matrix is said to be **full rank** if its rank is the largest possible, i.e., $\operatorname{rank}(A) = \min(m, n)$.

Definition 4.9. (Null Space, and Nullity) The **null space** of a matrix, $A \in \mathbb{F}^{m,n}$, denoted $\operatorname{null} A$, is the subset of $\mathbb{F}^m$ such that $Ax = \mathbf{0}$ for any $x \in \operatorname{null} A$, i.e.,

$$\operatorname{null} A := \{x | Ax = \mathbf{0}\}.$$

The **nullity** of A, denoted $\operatorname{nullity} A$ is the dimension of its null space, i.e. $\operatorname{nullity} A = \dim(\operatorname{null} A)$.

Theorem 4.1. (Rank-Nullity Theorem) For any matrix $A \in \mathbb{F}^{n,m}$, we have $\operatorname{rank} A + \operatorname{nullity} A = m$.

Theorem 4.2. (Sufficient Conditions for Diagonalizability) A matrix, A, is diagonalizable if and only if it has n distinct eigenvalues. If $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues, and $v_1, \ldots, v_n$ are the corresponding eigenvectors, then

$$A = P\Delta P^{-1},$$

where $\Delta = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalues and

$$P = \begin{bmatrix} | & & | \\ v_1 & \ldots & v_n \\ | & & | \end{bmatrix},$$

is a matrix of corresponding eigenvectors.

Proof. We begin by showing that if A has n distinct eigenvalues, the corresponding eigenvectors are linearly independent.

Let (λ_i, v_i) the i^{th} eigenvalue-eigenvector pair of the matrix A and assume $v_1, \ldots, v_n$ are linearly dependent. Find the smallest subset of $v_1, \ldots, v_n$ that are still linearly dependent. Without loss of generality, denote these vectors as $v_1, \ldots, v_j$. It follows that

$$\sum_{i=1}^{j} \alpha_i v_i = 0,$$

where at least two of $\alpha_1, \ldots, \alpha_j$ are not zero. If we multiply both sides of the equation by A, we have

$$\sum_{i=1}^{j} \alpha_i A v_i = \sum_{i=1}^{j} \alpha_i \lambda_i v_i = 0.$$

If instead, we multiply both sides of the equation by λ_1 we have

$$\sum_{i=1}^{j} \alpha_i \lambda_1 v_i = 0.$$

Equating the latter two equations yields

$$\sum_{i=1}^{j} \alpha_i \lambda_1 v_i = \sum_{i=1}^{j} \alpha_i \lambda_i v_i$$

$$\Rightarrow \sum_{i=2}^{j} \alpha_i(\lambda_1 - \lambda_i)v_i = 0.$$

Since the λ_i are distinct, $\lambda_1 - \lambda_i$ must be non-zero for all i. Moreover, at least one of $\alpha_2, \ldots, \alpha_j$ must also be non-zero. Thus, the above equation implies that $v_2, \ldots, v_j$ are linearly dependent. However, this is a contradiction since we assumed that $v_1, \ldots, v_j$ is the smallest linearly dependent set. It follows that $v_1, \ldots, v_n$ must be linearly independent.

We now show that if A has n linearly independent eigenvectors, it is diagonalizable. In particular, we will show that $A = P\Delta P^{-1}$, where

$$P = \begin{bmatrix} | & & | \\ v_1 & \ldots & v_n \\ | & & | \end{bmatrix} \text{ and } \Delta = \text{diag}(\lambda_1, \ldots, \lambda_n).$$

We have

$$AP = \begin{bmatrix} | & & | \\ Av_1 & \ldots & Av_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1 v_1 & \ldots & \lambda_n v_n \\ | & & | \end{bmatrix} = P\Delta.$$

Since the columns of P are linearly independent, P is invertible and we may write $A = P\Delta P^{-1}$. $\square$