- $\mathbf{P}_{231} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ This matrix shifts rows two and three up one and moves row one to the position of row three of the matrix it is applied on.
- $\mathbf{P}_{312} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ This matrix shifts rows one and two down one and moves row three to the row-one position of the matrix it is applied on.
- $\mathbf{P}_{321} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ This matrix swaps rows one and three of the matrix it is applied on.

It is important to note that there is a particularly fascinating property of permutation matrices that states that if we have a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and it is invertible, then there exists a permutation matrix that when applied to A will give us the LU factor of A. We can express this like so:

$$\mathbf{PA} = \mathbf{LU}$$

Vector spaces and subspaces

In this section, we will explore the concepts of vector spaces and subspaces. These are very important to our understanding of linear algebra. In fact, if we do not have an understanding of vector spaces and subspaces, we do not truly have an understanding of how to solve linear algebra problems.

Spaces

Vector spaces are one of the fundamental settings for linear algebra, and, as the name suggests, they are spaces where all vectors reside. We will denote the vector space with V.

The easiest way to think of dimensions is to count the number of elements in the column vector. Suppose we have $\mathbf{x} = (x_1, x_2, \cdots, x_7)$, then $\mathbf{x} \in \mathbb{R}^7$. $\mathbb{R}^1$ is a straight line, $\mathbb{R}^2$ is all the possible points in the *xy*-plane, and $\mathbb{R}^3$ is all the possible points in the *xyz*-plane—that is, 3-dimensional space, and so on.

The following are some of the rules for vector spaces:

- There exists in *V* an additive identity element such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in V$.
- For all $\mathbf{x} \in V$, there exists an additive inverse such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- For all $\mathbf{x} \in V$, there exists a multiplicative identity such that $1\mathbf{x} = \mathbf{x}$.
- Vectors are commutative, such that for all $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
- Vectors are associative, such that $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
- Vectors have distributivity, such that $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ and $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in V$ and for all $\alpha, \beta \in \mathbb{R}$.

A set of vectors is said to be linearly independent if $\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$, which implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

Another important concept for us to know is called **span**. The span of $\mathbf{v}_1, \cdots, \mathbf{v}_n \in V$ is the set of all linear combinations that can be made using the *n* vectors. Therefore, $\text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = \{\mathbf{v} \in V : \exists\alpha_1, \cdots, \alpha_n \mid \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{v}\}$ if the vectors are linearly independent and span *V* completely; then, the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are the basis of *V*.

Therefore, the dimension of *V* is the number of basis vectors we have, and we denote it *dimV*.

Subspaces

Subspaces are another very important concept that state that we can have one or many vector spaces inside another vector space. Let's suppose *V* is a vector space, and we have a subspace $S \subseteq V$. Then, *S* can only be a subspace if it follows the three rules, stated as follows:

- $\mathbf{0} \in S$
- $\mathbf{x}, \mathbf{y} \in S$ and $\mathbf{x} + \mathbf{y} \in S$, which implies that *S* is closed under addition
- $\mathbf{x} \in S$ and $\alpha \in \mathbb{R}$ so that $\alpha\mathbf{x} \in S$, which implies that *S* is closed under scalar multiplication

If $U, W \in V$, then their sum is $U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}$, where the result is also a subspace of *V*.

The dimension of the sum $U + W$ is as follows:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$$

Linear maps

A linear map is a function $T : V \to W$, where *V* and *W* are both vector spaces. They must satisfy the following criteria:

- $T(\mathbf{x}+\mathbf{y}) = T\mathbf{x} + T\mathbf{y}$, for all $\mathbf{x}, \mathbf{y} \in V$
- $T(\alpha\mathbf{x}) = \alpha T\mathbf{x}$, for all $\mathbf{x} \in V$ and $\alpha \in \mathbb{R}$

Linear maps tend to preserve the properties of vector spaces under addition and scalar multiplication. A linear map is called a **homomorphism of vector spaces;** however, if the homomorphism is invertible (where the inverse is a homomorphism), then we call the mapping an **isomorphism**.

When *V* and *W* are isomorphic, we denote this as $V \cong W$, and they both have the same algebraic structure.

If *V* and *W* are vector spaces in $\mathbb{R}^n$, and $\dim V = \dim W = n$, then it is called a **natural isomorphism**. We write this as follows:

$$\varphi : V \to W$$
$$\alpha_1 \mathbf{v}_1 + \alpha_n \mathbf{v}_n \mapsto \alpha_1 \mathbf{w}_1 + \alpha_n \mathbf{w}_n$$

Here, $\mathbf{v}_1, \cdots, \mathbf{v}_n$ and $\mathbf{w}_1, \cdots, \mathbf{w}_n$ are the bases of *V* and *W*. Using the preceding equation, we can see that $V \cong W$, which tells us that φ is an isomorphism.

Let's take the same vector spaces *V* and *W* as before, with bases $\mathbf{v}_1, \cdots, \mathbf{v}_n$ and $\mathbf{w}_1, \cdots, \mathbf{w}_m$ respectively. We know that $T : V \to W$ is a linear map, and the matrix *T* that has entries A_{ij}, where $i = 1, \cdots, m$ and $j = 1, \cdots, n$ can be defined as follows:

$$T\mathbf{v}_j = A_{1,j}\mathbf{w}_1 + \cdots + A_{m,j}\mathbf{w}_m.$$

From our knowledge of matrices, we should know that the j^{th} column of A contains Tv_j in the basis of *W*.

Thus, $\mathbf{A} \in \mathbb{R}^{m \times n}$ produces a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$, which we write as $T\mathbf{x} = \mathbf{A}\mathbf{x}$.

Image and kernel

When dealing with linear mappings, we will often encounter two important terms: the image and the kernel, both of which are vector subspaces with rather important properties.

The **kernel** (sometimes called the **null space**) is 0 (the zero vector) and is produced by a linear map, as follows:

$$\ker(T) = \{\mathbf{v} \in V \mid T\mathbf{v} = \mathbf{0}\}$$

And the **image** (sometimes called the **range**) of T is defined as follows:

$$\mathrm{Im}(T) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \text{ such that } T\mathbf{v} = \mathbf{w}\}.$$

V and *W* are also sometimes known as the **domain** and **codomain** of *T*.

It is best to think of the kernel as a linear mapping that maps the vectors $\mathbf{v} \in V$ to $\mathbf{0} \in W$. The image, however, is the set of all possible linear combinations of $\mathbf{v} \in V$ that can be mapped to the set of vectors $\mathbf{w} \in W$.

The **Rank-Nullity theorem** (sometimes referred to as the **fundamental theorem of linear mappings**) states that given two vector spaces *V* and *W* and a linear mapping $T : V \to W$, the following will remain true:

$$\dim(\ker(T)) + \dim(\mathrm{Im}(T)) = \dim(V).$$

Metric space and normed space

Metrics help define the concept of distance in Euclidean space (denoted by $\mathbb{E}^n$). Metric spaces, however, needn't always be vector spaces. We use them because they allow us to define limits for objects besides real numbers.

So far, we have been dealing with vectors, but what we don't yet know is how to calculate the length of a vector or the distance between two or more vectors, as well as the angle between two vectors, and thus the concept of orthogonality (perpendicularity). This is where Euclidean spaces come in handy. In fact, they are the fundamental space of geometry. This may seem rather trivial at the moment, but their importance will become more apparent to you as we get further on in the book.

This satisfies the rules for metrics, telling us that a normed space is also a metric space.

In general, for our purposes, we will only be concerned with four norms on $\mathbb{R}^n$, as follows:

- $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
- $\|\mathbf{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$
- $\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$ (this applies only if $p \geq 1$)

If you look carefully at the four norms, you can notice that the 1- and 2-norms are versions of the p-norm. The ∞-norm, however, is a limit of the p-norm, as p tends to infinity.

Using these definitions, we can define two vectors to be orthogonal if the following applies:

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Inner product space

An inner product on a vector space is a function $\langle v_1, v_2 \rangle : V \times V \to \mathbb{R}$, and satisfies the following rules:

- $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$
- $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ and $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$
- $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha \in \mathbb{R}$.

It is important to note that any inner product on the vector space creates a norm on the said vector space, which we see as follows:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

We can notice from these rules and definitions that all inner product spaces are also normed spaces, and therefore also metric spaces.

Another very important concept is orthogonality, which in a nutshell means that two vectors are perpendicular to each other (that is, they are at a right angle to each other) from Euclidean space.

Two vectors are orthogonal if their inner product is zero—that is, $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. As a shorthand for perpendicularity, we write $\mathbf{x} \perp \mathbf{y}$.

Additionally, if the two orthogonal vectors are of unit length—that is, $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, then they are called **orthonormal**.

In general, the inner product in $\mathbb{R}^n$ is as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}^{\mathrm{T}} \mathbf{y}$$

Matrix decompositions

Matrix decompositions are a set of methods that we use to describe matrices using more interpretable matrices and give us insight to the matrices' properties.

Determinant

Earlier, we got a quick glimpse of the determinant of a square 2x2 matrix when we wanted to determine whether a square matrix was invertible. The determinant is a very important concept in linear algebra and is used frequently in the solving of systems of linear equations.

Note: The determinant only exists when we have square matrices.

Notationally, the determinant is usually written as either $\det(\mathbf{A})$ or $|\mathbf{A}|$.

Let's take an arbitrary *n*×*n* matrix A, as follows:

Singular value decomposition

Singular Value Decomposition (**SVD**) is widely used in linear algebra and is known for its strength, particularly arising from the fact that every matrix has an SVD. It looks like this:

$$\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$$

For our purposes, let's suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$, and $\mathbf{V} \in \mathbb{R}^{n \times n}$, and that *U*, *V* are orthogonal matrices, whereas ∑ is a matrix that contains singular values (denoted by σ_i) of *A* along the diagonal.

∑ in the preceding equation looks like this:

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \end{bmatrix} \mathbf{V}^{\mathrm{T}}$$

We can also write the SVD like so:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}$$

Here, u_i, v_i are the column vectors of *U*, *V*.

Cholesky decomposition

As I'm sure you've figured out by now, there is more than one way to factorize a matrix, and there are special methods for special matrices.

The Cholesky decomposition is square root-like and works only on symmetric positive definite matrices.

This works by factorizing *A* into the form LL^T. Here, *L*, as before, is a lower triangular matrix.

Do develop some intuition. It looks like this:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} = \begin{bmatrix} l_{1,1} & 0 & \cdots & 0 \\ l_{2,1} & l_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n} \end{bmatrix} \begin{bmatrix} l_{1,1} & l_{1,2} & \cdots & l_{1,n} \\ 0 & l_{2,2} & \cdots & l_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_{n,n} \end{bmatrix}$$

However, here, L is called a **Cholesky factor**.

Let's take a look at the case where $\mathbf{A} \in \mathbb{R}^{3\times 3}$.

We know from the preceding matrix that $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$; therefore, we have the following:

$$\begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{2,1} & a_{2,2} & a_{3,2} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = \begin{bmatrix} l_{1,1} & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix} \begin{bmatrix} l_{1,1} & l_{2,1} & l_{3,1} \\ 0 & l_{2,2} & l_{3,2} \\ 0 & 0 & l_{3,3} \end{bmatrix}$$

Let's multiply the upper and lower triangular matrices on the right, as follows:

$$\mathbf{A} = \begin{bmatrix} l_{1,1}^2 & l_{2,1}l_{1,1} & l_{3,1}l_{1,1} \\ l_{2,1}l_{1,1} & l_{2,1}^2 l_{2,2}^2 & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} \\ l_{3,1}l_{1,1} & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} & l_{3,1}^2 + l_{3,2}^2 + l_{3,3}^2 \end{bmatrix}$$

Writing out A fully and equating it to our preceding matrix gives us the following:

$$\begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{2,1} & a_{2,2} & a_{3,2} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} = \begin{bmatrix} l_{1,1}^2 & l_{2,1}l_{1,1} & l_{3,1}l_{1,1} \\ l_{2,1}l_{1,1} & l_{2,1}^2 l_{2,2}^2 & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} \\ l_{3,1}l_{1,1} & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} & l_{3,1}^2 + l_{3,2}^2 + l_{3,3}^2 \end{bmatrix}$$

We can then compare, element-wise, the corresponding entries of A and LL^T and solve algebraically for $l_{i,j}$, as follows:

$$
\begin{aligned}
l_{1,1} &= \sqrt{a_{1,1}} \\
l_{2,1} &= \frac{1}{l_{1,1}} a_{2,1} \\
l_{2,2} &= \sqrt{a_{2,2} - l_{2,1}^2} \\
l_{3,1} &= \frac{1}{l_{1,1}} a_{3,1} \\
l_{3,2} &= \frac{1}{l_{2,2}} (a_{3,2} - l_{3,1} l_{2,1}) \\
l_{3,3} &= \sqrt{a_{3,3} - l_{3,1}^2 + l_{3,2}^2}
\end{aligned}
$$

We can repeat this process for any symmetric positive definite matrix, and compute the $l_{i,j}$ values given $a_{i,j}$.

Summary

With this, we conclude our chapter on linear algebra. So far, we have learned all the fundamental concepts of linear algebra, such as matrix multiplication and factorization, that will lead you on your way to gaining a deep understanding of how **deep neural networks** (**DNNs**) work and are designed, and what it is that makes them so powerful.

In the next chapter, we will be learning about calculus and will combine it with the concepts learned earlier on in this chapter to understand vector calculus.