Appendix B
Linear Algebra

B.1 VECTOR SPACES

In addition to constructing laws of composition we need our operations to be well behaved. Sets with this additional mathematical structure are vector spaces.

Definition B.1. *A real vector space V is a collection of elements of objects with two laws of composition. The first defines vector addition*

$$\begin{array}{ccc} V \times V & \to & V \\ (\mathbf{v}, \mathbf{w}) & \rightsquigarrow & \mathbf{v} + \mathbf{w} \end{array}$$

and the second defines scalar multiplication.

$$\begin{array}{ccc} \mathbb{R} \times V & \to & V \\ (c, \mathbf{v}) & \rightsquigarrow & c\mathbf{v} \end{array}$$

In addition the following properties must be satisfied:

- $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- $0 \in V$ such that $0 + \mathbf{v} = \mathbf{v} + 0$
- $\mathbf{u} + (-1)\mathbf{u} = 0$

- $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$, commutative law
- $(ab)\mathbf{v}=a(b\mathbf{v})$ associative law
- $(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$
- $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$
- $1\mathbf{u}=\mathbf{u}$

Note that multiplication of vectors plays no role in the definition of a vector space.

Example B.1. Real n-dimensional Cartesian space $\mathbb{R}^n$. Let $\mathbf{x}=(x_1,x_2,\dots,x_n)$ and $\mathbf{y}=(y_1,y_2,\dots,y_n)$ and $\alpha\in\mathbb{R}$. Define

$$\mathbf{x}+\mathbf{y}=(x_1+y_1,x_2+y_2,\dots,x_n+y_n)$$

$$\alpha\mathbf{x}=(\alpha x_1,\alpha x_2,\dots,\alpha x_n)$$

Example B.2. Consider the set of all real valued functions T. If $f,g\in T$ we define addition

$$(f+g)(x)=f(x)+g(x)$$

and scalar multiplication

$$(\alpha f)(x)=\alpha f(x).$$

With these definitions T forms a vector space of functions.

Definition B.2. *A* **linear combination** *of vectors* $\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_n\}$ *is any vector of the form*

$$\mathbf{w}=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_n\mathbf{v}_n.$$

Definition B.3. *A collection of vectors* $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ *is called* **linearly independent** *if there is no expression of the form*

$$0=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$$

except for the trivial case that $c_i=0$ *for all* i. *A collection of vectors which is not linearly independent is called* **linearly dependent**.

Definition B.4. *A family of vectors which is linearly independent and spans* V *is called a* **basis**. *By convention we consider a basis as an ordered set of vectors.*

Definition B.5. *Span(S) denotes the subspace which the set of vectors* S *spans.*

Proposition B.1. *Let* L *be any linearly independent family of vectors contained in a vector space* V. *The family* L' *obtained by adding* $\mathbf{v}\in V$ *to* L, *i.e.,*

$L' = L \cup \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v}$ is not in the subspace spanned by L.

Proposition B.2. *Any linearly independent family can be extended by adding elements to get a basis.*

Definition B.6. *The* **dimension** *of a finite dimensional vector space V, written $dimV$, is the number of vectors in the basis.*

The dimension of the vector space according to the above definition will always be a whole number. Note that there are several alternative measures of dimensionality such as the Hausdorff dimension, the topological dimension and the information dimension.

B.2 MAPPINGS

Definition B.7. *Let S, S' be two sets. A* **mapping** *T from S to S' is an association which relates every element of S to an element of S'. Formally,*

$$\begin{array}{rcl} T : S & \rightarrow & S' \\ u & \rightsquigarrow & T(u) \end{array}$$

(If $S' = R$ we refer to the mapping as a function). $T(u)$ is called the image of u under T and $T(S)$ is called the image of S under T.

Example B.3. Let $\mathbf{u}$ be a fixed vector in V. Define the mapping $T_{\mathbf{u}}$ as

$$\begin{array}{rcl} T_{\mathbf{u}} : V & \rightarrow & V \\ T_{\mathbf{u}}(\mathbf{v}) & = & \mathbf{v} + \mathbf{u} \end{array}$$

$T_{\mathbf{u}}$ is referred to as a **translation** of $\mathbf{v}$ by $\mathbf{u}$.

The connection between $\mathbb{R}^n$ and an n-dimensional vector space is established by choosing a coordinate system. To make this more rigorous we introduce a fundamental concept.

Definition B.8. *An* **isomorphism** *ϕ from a vector space V to a vector space V' is a bijective map $\phi : V \rightarrow V'$ such that*

$$\phi(\mathbf{v} + \mathbf{v}') = \phi(\mathbf{v}) + \phi(\mathbf{v}')$$

$$\phi(c\mathbf{v}) = c\phi(\mathbf{v}).$$

In other words, the map is compatible with the laws of composition.

Definition B.9. *Two vector spaces V, V' are said to be* **isomorphic** *if there is a $1-1$ correspondence between the vectors $\mathbf{v} \in V$ and $\mathbf{v}' \in V'$.*

Isomorphism is a very strong relationship. In fact if two vector spaces are isomorphic then they are essentially carbon copies of each other in a mathematical sense. We will use this idea later when we want to establish the equivalence of different representations of pattern vectors.

Theorem B.1. *Every n-dimensional vector space U over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$.*

Proof B.2.1. *Given a basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of U we write*

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n.$$

Take ϕ to be the map of $\mathbf{u} \in U$ onto its unique coordinates (n-tuple)

$$\phi(\mathbf{u}) = (a_1, a_2, \ldots, a_n) = \mathbf{a} \in \mathbb{R}^n.$$

Also, if $\mathbf{u}' \in U$ then we can write

$$\mathbf{u}' = a_1'\mathbf{u}_1 + a_2'\mathbf{u}_2 + \cdots + a_n'\mathbf{u}_n.$$

Then

$$\alpha\mathbf{u} + \beta\mathbf{u}' = (\alpha a_1 + \beta a_1')\mathbf{u}_1 + \cdots + (\alpha a_n + \beta a_n')\mathbf{u}_n.$$

Thus we have the 1-1 correspondence

$$\text{(coordinate free)} \quad \mathbf{u} \leftrightarrow \mathbf{a} = (a_1, \ldots, a_n) \ \text{w.r.t. the basis } \mathcal{B} \tag{B.1}$$

B.3 INNER PRODUCT SPACES

While the vector space provided the necessary framework within which to manipulate patterns we need additional structure to relate patterns. In the following we will assume that $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are elements of a vector space V and $\alpha \in \mathbb{R}$.

A vector space V is called an **inner product space** if there is defined a law of composition which maps any two elements of V onto the real line,i.e.,

$$V \times V \to \mathbb{R}$$

$$(\mathbf{u}, \mathbf{v}) \rightsquigarrow \mathbb{R}$$

with the following properties:

- $(\mathbf{u} + \mathbf{v}, \mathbf{w}) = (\mathbf{u}, \mathbf{w}) + (\mathbf{v}, \mathbf{w})$. Linearity.
- $(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})$. Symmetry.
- $(\alpha\mathbf{u}, \mathbf{v}) = \alpha(\mathbf{u}, \mathbf{v})$. Homogeneity.
- $(\mathbf{u}, \mathbf{u}) \geq 0$ and $(\mathbf{u}, \mathbf{u}) = 0$ iff $\mathbf{u} = \mathbf{0}$. Positivity.

One of the most fundamental relationships between vectors is given by their inner product being 0.

Definition B.10. *Two vectors* $\mathbf{u}, \mathbf{v}$ *are said to be orthogonal if* $(\mathbf{u}, \mathbf{v}) = 0$.

The quantity $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}$ is called the inner product of $\mathbf{u}$ and $\mathbf{v}$. For a complex inner product space $(\mathbf{u}, \mathbf{v}) \in \mathbb{C}$ and we replace the symmetry condition above with $(\mathbf{u}, \mathbf{v}) = \overline{(\mathbf{v}, \mathbf{u})}$, an Hermitian symmetry condition.

Example B.4. $V = \mathbb{R}^n$ the weighted Euclidean inner product is given by

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} w_i u_i v_i$$

with each $w_i \geq 0$.

Example B.5. $V = \mathbb{C}^n$ the weighted complex Euclidean inner product is given by

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} w_i u_i \overline{v}_i$$

with each $w_i \geq 0$.

B.4 VECTOR AND MATRIX NORMS

The *norm* of a vector is a unary operator which maps any element of a vector space V onto the real line and satisfies the properties

- $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$. Triangle inequality.
- $||\alpha \mathbf{u}|| = |\alpha| ||\mathbf{u}||$.
- $||\mathbf{u}|| \geq 0$. Equality iff $\mathbf{u} = 0$.

The following theorem permits a connection between any inner product space and a normed vector space.

Theorem B.2. *If* V *is an inner product space, the equation*

$$||\mathbf{u}|| = \sqrt{(\mathbf{u}, \mathbf{u})}$$

defines a norm in V *and* V *is referred to as a normed vector space.*

Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$. Examples of vector norms are

- The l_1 norm defined as

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_n| \tag{B.2}$$

- The l_2 norm, or standard Euclidean norm, defined as

$$||\mathbf{x}||_2 = \sum_{i=1}^{n} \sqrt{\mathbf{x}^T \mathbf{x}} \tag{B.3}$$

- The l_2 norm, or standard Euclidean norm, defined as

$$||\mathbf{x}||_\infty = \max_{1 \leq i \leq n} |x_i| \tag{B.4}$$

- The weighted norm defined as

$$||\mathbf{x}||_W = ||W\mathbf{x}|| \tag{B.5}$$

  where W is any non-singular matrix.

Matrix norms are defined in the same fashion as vector norms, i.e., a matrix norm must satisfy

- $||A + B|| \leq ||A|| + ||B||$. Triangle inequality.
- $||\alpha A|| = |\alpha|||A||$.
- $||\mathbf{A}|| \geq 0$. Equality iff $A = 0$.

where A and B are matrices.

A widely used example of a matrix norm is the Frobenius norm which is defined as

$$||A||_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_i j|^2. \tag{B.6}$$

A special class of matrix norms are referred to as *induced matrix norms.* They inherit there measure from vector norms. A matrix norm is induced by a vector norm if

$$||A|| = \sup_{x \neq 0} \frac{||A\mathbf{x}||}{||\mathbf{x}||}$$

In other words, $||A||$ is the smallest number α such that

$$\frac{||A\mathbf{x}||}{||\mathbf{x}||} \leq \alpha$$

for all $\mathbf{x}$. The induced matrix norm is a measure of how much the matrix A expands a vector, i.e., it is defined by its action.

Induced matrix norms may be induced by different vector norms. The nature of the norm is identified by the appropriate subscript, i.e., $||A||_2$ for the induced 2-norm.

Example B.6. Let Σ be a diagonal matrix with elements $\sigma^{(i)}$ on the diagonal. Then

$$||\Sigma||_2 = \max_{||\mathbf{x}||=1} ||\Sigma\mathbf{x}||_2$$

hence

$$||\Sigma||_2 = \max_i |\sigma^{(i)}||$$

Orthogonal matrices also preserve 2-norms (and F-norms). If Q, P orthogonal then

$$||QAP|| = ||A||$$

This is not true in general, e.g., it is false for 1-norms.

It follows from these two statements that the 2-norm of a matrix is given by the largest singular value in the SVD. If $A = U\Sigma V^T$, then $||A||_2 = ||\Sigma||_2$ so

$$||A||_2 = \sigma^{(1)}$$

B.5 METRIC SPACES

Definition B.11. *A* **metric space** *is a collection of elements and a law of composition which defines for every pair of elements belonging to the set a distance function which satisfies the properties*

- $d(\mathbf{u}, \mathbf{v}) \geq 0$ *with equality iff* $\mathbf{u} = \mathbf{v}$.
- $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
- $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

Example B.7.

$$d_N(\mathbf{u}, \mathbf{v}) = \frac{d(\mathbf{u}, \mathbf{v})}{1 + d(\mathbf{u}, \mathbf{v})}.$$

If $d(\mathbf{u}, \mathbf{v})$ is a metric show that $d_N(\mathbf{u}, \mathbf{v})$ also defines a metric. What is the range of possible values of $d_N(\mathbf{u}, \mathbf{v})$.

Solution: To show that d_N is a metric we must demonstrate that all the properties in the definition are satisfied. First note that $d_N(u, v) = 0$ iff $d(\mathbf{u}, \mathbf{v}) = 0$ therefore $d_N(\mathbf{u}, \mathbf{v}) = 0$ iff $\mathbf{u} = \mathbf{v}$. If $\mathbf{u} \neq \mathbf{v}$ then $d(\mathbf{u}, \mathbf{v}) > 0$, i.e., d_N behaves like the function $f(x) = \frac{x}{1+x}$. We have $f'(x) = (1 + x)^{-2}$ which is monotonically increasing. Hence, $d_N(\mathbf{u}, \mathbf{v})$ is always positive. To show the second property we have

$$d_N(\mathbf{v}, \mathbf{u}) = \frac{d(\mathbf{v}, \mathbf{u})}{1 + d(\mathbf{v}, \mathbf{u})} = \frac{d(\mathbf{u}, \mathbf{v})}{1 + d(\mathbf{u}, \mathbf{v})} = d_N(\mathbf{u}, \mathbf{v}).$$

The triangle inequality for $d_N(\mathbf{u}, \mathbf{v})$ is a straightforward consequence of the fact that $d(\mathbf{u}, \mathbf{v})$ satisfies the triangle inequality.

To find the maximum value of d_N (we have already shown that the minimum is 0) we observe that d_N increases monotonically with d. Hence

$$\max d_N = \lim_{d\to\infty} \frac{d}{1+d} = 1$$

I.e., $d_N(\mathbf{u},\mathbf{v}) \in [0,\infty)$ for all $\mathbf{u},\mathbf{v}$.

The similarity of the definitions of metric and norm should be apparent and in fact it can easily be shown that every normed linear space is a metric space if we define the metric as

$$d(\mathbf{u},\mathbf{v}) = ||\mathbf{u}-\mathbf{v}||.$$

We now present several examples of metrics which are used in the literature.

Example B.8. The Minkowski metric is given by

$$d_{M(\lambda)}(\mathbf{u},\mathbf{v}) = (\sum_{i=1}^{n} |u_i - v_i|^{\lambda})^{1/\lambda}.$$

$d_{M(1)}$ is often referred to as the *city block* metric.

Example B.9. The Tanimoto similarity measure is defined on a normed inner product space and is given by

$$d(\mathbf{u},\mathbf{v}) = \frac{(\mathbf{u},\mathbf{v})}{||\mathbf{u}||^2 + ||\mathbf{v}||^2 - (\mathbf{u},\mathbf{v})}.$$

Example B.10. The discrete metric requires only a set of objects and identifies the distance of all non-identical objects as having the value 1, specifically

$$d(\mathbf{u},\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{u} = \mathbf{v} \\ 1 & \text{if } \mathbf{u} \neq \mathbf{v} \end{cases}$$

Example B.11. The Hamming distance between two members of a set is defined as the number of elements of each member which do no match. For instance if $\mathbf{u} = (1,0,1,1,1,0), \mathbf{v} = (1,1,1,1,0,0)$ then $d_H = 2$. As another example if $\mathbf{u} = (c,a,t), \mathbf{v} = (h,a,t)$ then $d_H = 1$.

Example B.12. The *Mahalanobis* distance

$$d_M(\mathbf{u},\mathbf{v}) = ((\mathbf{u}-\mathbf{v})^T \mathbf{C}^{-1} (\mathbf{u}-\mathbf{v}))^{1/2}$$

where $\mathbf{C} = \mathbf{u}\mathbf{v}^T$ is the covariance matrix of $\mathbf{u}$ and $\mathbf{v}$ is an optimal distance for vectors corrupted by normally distributed noise.