LINEAR MAPPINGS

SYNOPSIS. Vectors have been introduced to represent complicated objects, whose description requires m numbers, and the procedure of linear combination allows construction of new vectors. Alternative insights into some object might be obtained by transformation of vectors. Of all possible transformations, those that are compatible with linear combinations are of special interest. It turns out that matrices are not only important in organizing collections of vectors, but also to represent such transformations, referred to as *linear mappings*.

1. Functions

1.1. Relations

The previous chapter focused on mathematical expression of the concept of *quantification*, the act of associating human observation with measurements, as a first step of scientific inquiry. Consideration of different types of quantities led to various types of numbers, vectors as groupings of numbers, and matrices as groupings of vectors. Symbols were introduced for these quantities along with some intial rules for manipulating such objects, laying the foundation for an *algebra* of vectors and matrices. Science seeks to not only observe, but to also explain, which now leads to additional operations for working with vectors and matrices that will define the framework of *linear algebra*.

Explanations within scientific inquiry are formulated as hypotheses, from which predictions are derived and tested. A widely applied mathematical transcription of this process is to organize hypotheses and predictions as two sets X and Y, and then construct another set R of all of the instances in which an element of X is associated with an element in Y. The set of all possible instances of $x \in X$ and $y \in Y$, is the *Cartesian product* of X with Y, denoted as $X \times Y = \{(x, y) | x \in X, y \in Y\}$, a construct already encountered in the definition of the real 2-space $\mathscr{R}_2 = (\mathbb{R}^2, \mathbb{R}, +, \cdot)$ where $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Typically, not all possible tuples $(x, y) \in X \times Y$ are relevant leading to the following definition.

DEFINITION. (RELATION) . *A relation R between two sets X, Y is a subset of the Cartesian product $X \times Y$, $R \subseteq X \times Y$.*

The key concept is that of *associating an input $x \in X$ with an output $y \in Y$*. Inverting the approach and associating an output to an input is also useful, leading to the definition of an *inverse relation* as $R^{-1} \subseteq Y \times X$, $R^{-1} = \{(y, x) | (x, y) \in R\}$. Note that an inverse exists for any relation, and the inverse of an inverse is the original relation, $(R^{-1})^{-1} = R$. From the above, a relation is a triplet (a tuple with three elements), (X, Y, R), that will often be referred to by just its last member R.

Homogeneous relations. Many types of relations are defined in mathematics and encountered in linear algebra, and establishing properties of specific relations is an important task within data science. A commonly encountered type of relationship is from a set onto itself, known as a *homogeneous* relation. For homogeneous relations $H \subseteq A \times A$, it is common to replace the set membership notation $(a, b) \in H$ to state that $a \in A$ is in relationship H with $b \in A$, with a binary operator notation $a \overset{H}{\sim} b$. Familiar examples include the equality and less than relationships between reals, $E, L \subseteq \mathbb{R} \times \mathbb{R}$, in which $(a, b) \in E$ is replaced by $a = b$, and $(a, b) \in L$ is replaced by $a < b$. The equality relationship is its own inverse, and the inverse of the less than relationship is the greater than relation $G \subseteq \mathbb{R} \times \mathbb{R}$, $G = L^{-1}$, $a < b \Longrightarrow b > a$.

1.2. Functions

Functions between sets X and Y are a specific type of relationship that often arise in science. For a given input $x \in X$, theories that predict a *single* possible output $y \in Y$ are of particular scientific interest.

DEFINITION. (FUNCTION) . *A function from set X to set Y is a relation $F \subseteq X \times Y$, that associates to $x \in X$ a single $y \in Y$.*

The above intuitive definition can be transcribed in precise mathematical terms as $F \subseteq X \times Y$ is a *function* if $(x, y) \in F$ and $(x, z) \in F$ implies $y = z$. Since it's a particular kind of relation, a function is a triplet of sets (X, Y, F), but with a special, common notation to denote the triplet by $f : X \to Y$, with $F = \{(x, f(x)) | x \in X, f(x) \in Y\}$ and the property that $(x, y) \in F \Rightarrow y = f(x)$. The set X is the *domain* and the set Y is the *codomain* of the function f. The value from the domain $x \in X$ is the *argument* of the function associated with the *function value* $y = f(x)$. The function value y is said to be *returned* by evaluation $y = f(x)$.

Whereas all relations can be inverted, and inversion of a function defines a new relation, but which might not itself be a function. For example the relation $S^{-1} = \{(\alpha, a), (\beta, a), (\gamma, a)\}$ is a function, but its inverse $(S^{-1})^{-1} = S$ is not.

Familiar functions include:

- the trigonometric functions $\cos : \mathbb{R} \to [-1, 1]$, $\sin : \mathbb{R} \to [-1, 1]$ that for argument $\theta \in \mathbb{R}$ return the function values $\cos(\theta), \sin(\theta)$ giving the Cartesian coordinates $(x, y) \in \mathbb{R}^2$ of a point on the unit circle at angular extent θ from the x-axis;

- the exponential and logarithm functions $\exp : \mathbb{R} \to \mathbb{R}$, $\log : (0, \infty) \to \mathbb{R}$, as well as power and logarithm functions in some other base a;

- polynomial functions $p_n : \mathbb{R} \to \mathbb{R}$, defined by a succession of additions and multiplications

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = \sum_{i=0}^{n} a_i x^i = ((a_n x + a_{n-1}) x + \cdots + a_1) x + a_0.$$

Simple functions such as sin, cos, exp, log, are predefined in Julia, and can be applied to each component of a vector argument by broadcasting, denoted by a period in front of the paranteses enclosing the argument.

```
∴ θ=π; [sin(θ) cos(θ) exp(θ) log(θ)]
```

$$[\; 0.0 \;\; -1.0 \;\; 23.140692632779267 \;\; 1.1447298858494002 \;] \tag{1}$$

```
∴ θ=0:π/6:π; short(x)=round(x,digits=6); short.(sin.(θ))'
```

$$[\; 0.0 \;\; 0.5 \;\; 0.866025 \;\; 1.0 \;\; 0.866025 \;\; 0.5 \;\; 0.0 \;] \tag{2}$$

```
∴ short.(log2.(1:8))'
```

$$[\; 0.0 \;\; 1.0 \;\; 1.584963 \;\; 2.0 \;\; 2.321928 \;\; 2.584963 \;\; 2.807355 \;\; 3.0 \;] \tag{3}$$

```
∴
```

A construct that will be often used is to interpret a vector within E_m as a function, since $\boldsymbol{v} \in \mathbb{R}^m$ with components $\boldsymbol{v} = [\ v_1 \ \ v_2 \ \ \ldots \ \ v_m\ ]^T$ also defines a function $v: \{1, 2, \ldots, m\} \to \mathbb{R}$, with values $v(i) = v_i$. As the number of components grows the function v can provide better approximations of some continuous function $f \in \mathscr{C}^0(\mathbb{R})$ through the function values $v_i = v(i) = f(x_i)$ at distinct sample points $x_1, x_2, \ldots, x_m$.

The above function examples are all defined on a domain of scalars or naturals and returned scalar values. Within linear algebra the particular interest is on functions defined on sets of vectors from some vector space $\mathscr{V} = (V, S, +, \cdot)$ that return either scalars $f: V \to S$, or vectors from some other vector space $\mathscr{W} = (W, S, +, \cdot)$, $\boldsymbol{g}: V \to W$. The codomain of a vector-valued function might be the same set of vectors as its domain, $\boldsymbol{h}: V \to V$. The fundamental operation within linear algebra is the linear combination $a\boldsymbol{u} + b\boldsymbol{v}$ with $a, b \in S$, $\boldsymbol{u}, \boldsymbol{v} \in V$. A key aspect is to characterize how a function behaves when given a linear combination as its argument, for instance $f(a\boldsymbol{u} + b\boldsymbol{v})$ or $\boldsymbol{g}(a\boldsymbol{u} + b\boldsymbol{v})$.

1.3. Linear functionals

Consider first the case of a function defined on a set of vectors that returns a scalar value. These can be interpreted as labels attached to a vector, and are very often encountered in applications from natural phenomena or data analysis.

DEFINITION. (FUNCTIONAL) . *A functional on vector space $\mathscr{V} = (V, S, +, \cdot)$ is a function from the set of vectors V to the set of scalars S of the vector space $\mathscr{V}$.*

DEFINITION. (LINEAR FUNCTIONAL) . *The functional $f: V \to S$ on vector space $\mathscr{V} = (V, S, +, \cdot)$ is a linear functional if for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ and any two scalars a, b*

$$f(a\boldsymbol{u} + b\boldsymbol{v}) = af(\boldsymbol{u}) + bf(\boldsymbol{v}). \tag{4}$$

1.4. Linear mappings

Consider now functions $\boldsymbol{f}: V \to W$ from vector space $\mathscr{V} = (V, S, +, \cdot)$ to another vector space $\mathscr{W} = (W, T, +, \cdot)$. As before, the action of such functions on linear combinations is of special interest.

DEFINITION. (LINEAR MAPPING) . *A function $\boldsymbol{f}: V \to W$, from vector space $\mathscr{V} = (V, S, +, \cdot)$ to vector space $\mathscr{W} = (W, S, +, \cdot)$ is called a linear mapping if for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ and any two scalars $a, b \in S$*

$$\boldsymbol{f}(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{f}(\boldsymbol{u}) + b\boldsymbol{f}(\boldsymbol{v}). \tag{5}$$

The image of a linear combination $a\boldsymbol{u} + b\boldsymbol{v}$ through a linear mapping is another linear combination $a\boldsymbol{f}(\boldsymbol{u}) + b\boldsymbol{f}(\boldsymbol{v})$, and linear mappings are said to preserve the structure of a vector space.

Note that $f: \mathbb{R} \to \mathbb{R}$ defined as $y = f(x) = ax + b$ represents a line in the (x, y)-plane, but is not a linear mapping for $b \neq 0$ since

$$f(x + z) = a(x + z) + b \neq ax + b + az + b = f(x) + f(z).$$

Matrix-vector multiplication has been introduced as a concise way to specify a linear combination

$$f(x) = Ax = x_1 a_1 + \cdots + x_n a_n,$$

with $a_1, \ldots, a_n$ the columns of the matrix, $A = [\ a_1\ \ a_2\ \ \ldots\ \ a_n\ ]$. This is a linear mapping between the real spaces $\mathscr{R}_m, \mathscr{R}_n, f:\mathbb{R}^m \to \mathbb{R}^n$, and indeed any linear mapping between real spaces can be given as a matrix-vector product. Consider some $x \in \mathbb{R}^m$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_m \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1 e_1 + x_2 e_2 + \cdots + x_m e_m.$$

Applying the linear mapping f to x leads to

$$f(x) = f(x_1 e_1 + x_2 e_2 + \cdots + x_m e_m) = x_1 f(e_1) + x_2 f(e_2) + \cdots + x_m f(e_m).$$

The matrix A with columns $a_1 = f(e_1), \ldots, a_m = f(e_m)$ now allows finding

$$f(x) = Ax,$$

through a matrix-vector multiplication for any input vector x. The matrix A thus defined is a representation of the linear mapping f. As will be shown later, it is not the only possible representation.

2. Measurements

Vectors within the real space $\mathscr{R}_m$ can be completely specified by m real numbers, and m is large in many realistic applications. The task of describing the elements of a vector space $\mathscr{V} = (V, S, +, \cdot)$ by simpler means arises. Within data science this leads to *classification problems* in accordance with some relevant criteria, and one of the simplest classifications is to attach a scalar label to a vector. Commonly encountered labels include the magnitude of a vector or its orientation with respect to another vector.

2.1. Norms

The above observations lead to the mathematical concept of a *norm* as a tool to evaluate vector magnitude. Recall that a vector space is specified by two sets and two operations, $\mathscr{V} = (V, S, +, \cdot)$, and the behavior of a norm with respect to each of these components must be defined. The desired behavior includes the following properties and formal definition.

   Unique value. The magnitude of a vector $v \in V$ should be a unique scalar, requiring the definition of a function. The scalar could have irrational values and should allow ordering of vectors by size, so the function should be from V to $\mathbb{R}$, $f: V \to \mathbb{R}$. On the real line the point at coordinate x is at distance $|x|$ from the origin, and to mimic this usage the norm of $v \in V$ is denoted as $\|v\|$, leading to the definition of a function $\|\ \|: V \to \mathbb{R}_+$, $\mathbb{R}_+ = \{a \,|\, a \in \mathbb{R}, a \geqslant 0\}$.

Null vector case. Provision must be made for the only distinguished element of V, the null vector $\mathbf{0}$. It is natural to associate the null vector with the null scalar element, $\|\mathbf{0}\| = 0$. A crucial additional property is also imposed namely that the null vector is the *only* vector whose norm is zero, $\|\boldsymbol{v}\| = 0 \Rightarrow \boldsymbol{v} = \mathbf{0}$. From knowledge of a single scalar value, an entire vector can be determined. This property arises at key junctures in linear algebra, notably in providing a link to mathematical analysis, and is needed to establish the fundamental theorem of linear algbera or the singular value decomposition encountered later.

Scaling. Transfer of the scaling operation $\boldsymbol{v} = a\boldsymbol{u}$ property leads to imposing $\|\boldsymbol{v}\| = |a| \|\boldsymbol{u}\|$. This property ensures commensurability of vectors, meaning that the magnitude of vector $\boldsymbol{v}$ can be expressed as a multiple of some standard vector magnitude $\|\boldsymbol{u}\|$.

Vector addition. Position vectors from the origin to coordinates $x, y > 0$ on the real line can be added and $|x + y| = |x| + |y|$. If however the position vectors point in different directions, $x > 0$, $y < 0$, then $|x + y| < |x| + |y|$. For a general vector space the analogous property is known as the *triangle inequality*, $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ for $\boldsymbol{u}, \boldsymbol{v} \in V$.

DEFINITION. (NORM). *A norm on the vector space $\mathscr{V} = (V, S, +, \cdot)$ is a function $\| \, \|: V \to \mathbb{R}_+$ that for $\boldsymbol{u}, \boldsymbol{v} \in V$, $a \in S$ satisfies:*

1. $\|\boldsymbol{v}\| = 0 \Rightarrow \boldsymbol{v} = \mathbf{0}$;

2. $\|a\boldsymbol{u}\| = |a| \|\boldsymbol{u}\|$;

3. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.

A commonly encountered norm of $\boldsymbol{v} \in \mathbb{R}^m$ is the Euclidean norm

$$\|\boldsymbol{v}\| = \sqrt{v_1^2 + \cdots + v_m^2} = \left(v_1^2 + \cdots + v_m^2\right)^{1/2} = \left(\sum_{j=1}^{m} v_j^2\right)^{1/2},$$

useful in many physics applications. The form of the above norm, square root of sum of squares of components, can be generalized to obtain other useful norms.

DEFINITION. (p-NORM IN $\mathscr{R}_m$). *The p-norm on the real vector space $\mathscr{R}_m = (\mathbb{R}^m, \mathbb{R}, +, \cdot)$ for $p \geq 1$ is the function $\| \, \|_p: V \to \mathbb{R}_+$ with values $\|\boldsymbol{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_m|^p\right)^{1/p}$, or*

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p} \text{ for } \boldsymbol{x} \in \mathbb{R}^m. \tag{6}$$

Note that the Euclidean norm corresponds to $p = 2$, and is often called the 2-norm. Denote by x_i the largest component in absolute value of $\boldsymbol{x} \in \mathbb{R}^m$. As p increases, $|x_i|^p$ becomes dominant with respect to all other terms in the sum suggesting the definition of an inf-norm by

$$\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq m} |x_i|.$$

This also works for vectors with equal components, since the fact that the number of components is finite while $p \to \infty$ can be used as exemplified for $\boldsymbol{x} = [\, a \;\; a \;\; \ldots \;\; a \,]^T$, by $\|\boldsymbol{x}\|_p = \left(m |a|^p\right)^{1/p} = m^{1/p} |a|$, with $m^{1/p} \to 1$.
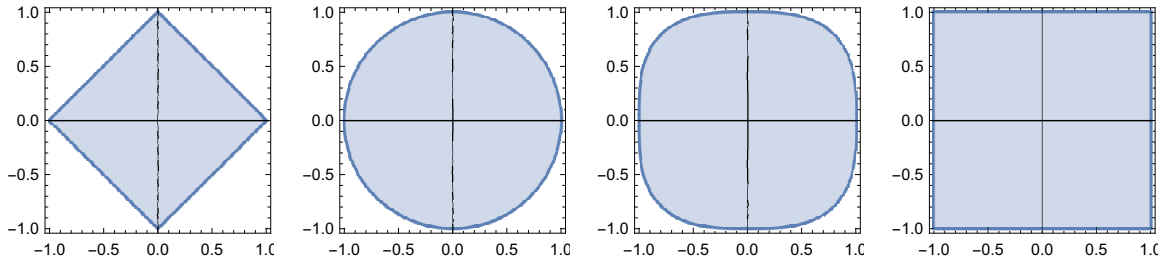
Figure 1. Regions within $\mathbb{R}^2$ for which $\|x\|_p \leq 1$, for $p = 1, 2, 3, \infty$.

Vector norms arise very often in applications, especially in data science since they can be used to classify data, and are implemented in software systems such as Julia in which the norm function with a single argument computes the most commonly encountered norm, the 2-norm. If a second argument p is specified the p-norm is computed.

```
∴ x=[1; 1; 1]; [norm(x) sqrt(3)]
```

$$[\ 1.7320508075688772 \quad 1.7320508075688772\ ] \tag{7}$$

```
∴ m=9; x=ones(m,1); [norm(x) sqrt(m)]
```

$$[\ 3.0 \quad 3.0\ ] \tag{8}$$

```
∴ m=4; x=ones(m,1); [norm(x,1) m]
```

$$[\ 4.0 \quad 4.0\ ] \tag{9}$$

```
∴
```

2.2. Inner product

Norms are functionals that define what is meant by the size of a vector, but are not linear. Even in the simplest case of the real line, the linearity relation $|x + y| = |x| + |y|$ is not verified for $x > 0$, $y < 0$. Nor do norms characterize the familiar geometric concept of orientation of a vector. A particularly important orientation from Euclidean geometry is orthogonality between two vectors. Another function is required, one that would take two vector arguments to enable characterizing their relative orientation. It would return a scalar, hence $s: V \times V \to S$, with S often chosen as the set of real numbers.

DEFINITION. (INNER PRODUCT) . *A real inner product in the vector space $\mathscr{V} = (V, \mathbb{R}, +, \cdot)$ is a function $s: V \times V \to \mathbb{R}$ with properties*

> **Symmetry.** *For any $a, x \in V$, $s(a, x) = s(x, a)$.*

> **Linearity in second argument.** *For any $a, x, y \in V$, $\alpha, \beta \in \mathbb{R}$, $s(a, \alpha x + \beta y) = \alpha s(a, x) + \beta s(a, y)$.*

> **Positive definiteness.** *For any $x \in V \backslash \{0\}$, $s(x, x) > 0$.*

A commonly encountered inner product is the dot product of two vectors $a, x \in \mathbb{R}^m$

$$a \cdot x = a_1 x_1 + \cdots + a_m x_m.$$

Using the convention of representing a, x as column vectors, the dot product is also expressed as

$$a^T x = \left[\begin{array}{cccc} a_1 & a_2 & \dots & a_m \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array}\right],$$

and is therefore a matrix multiplication between $a^T \in \mathbb{R}^{1 \times m}$ and $x \in \mathbb{R}^{m \times 1}$ resulting in a scalar, also referred to as a scalar product. Inner products also provide a procedure to evaluate geometrical quantities and relationships.

Vector norm. The square of the 2-norm of $x \in \mathbb{R}^m$ is given as

$$s(x, x) = x^T x = \|x\|_2^2.$$

In general, the square root of $s(x, x)$ satisfies the properties of a norm, and is called the norm induced by an inner product

$$\|x\| = s(x, x)^{1/2}.$$

A real space together with the scalar product $s(x, y) = x^T y$ and induced norm $\|x\| = s(x, x)^{1/2}$ defines an Euclidean vector space $\mathscr{E}_m$.

Orientation. In $\mathscr{E}_2$ the point specified by polar coordinates (r, θ) has the Cartesian coordinates $x_1 = r \cos \theta$, $x_2 = r \sin \theta$, and position vector $x = \left[\begin{array}{cc} x_1 & x_2 \end{array}\right]^T$. The inner product

$$e_1^T x = \left[\begin{array}{cc} 1 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = 1 \cdot x_1 + 0 \cdot x_2 = r \cos \theta,$$

is seen to contain information on the relative orientation of x with respect to e_1. In general, the angle θ between two vectors x, y with any vector space with a scalar product can be defined by

$$\cos \theta = \frac{s(x, y)}{[s(x, x) \, s(y, y)]^{1/2}} = \frac{s(x, y)}{\|x\| \|y\|},$$

which becomes

$$\cos \theta = \frac{x^T y}{\|x\| \|y\|},$$

in a Euclidean space, $x, y \in \mathbb{R}^m$.

Orthogonality. In $\mathscr{E}_2$ two vectors are orthogonal if the angle between them is such that $\cos \theta = 0$, and this can be extended to an arbitrary vector space $\mathscr{V} = (V, \mathbb{R}, +, \cdot)$ with a scalar product by stating that $x, y \in V$ are orthogonal if $s(x, y) = 0$. In $\mathscr{E}_m$ vectors $x, y \in \mathbb{R}^m$ are orthogonal if $x^T y = 0$.

3. Linear mapping matrices

3.1. Common geometric transformations

Several geometric transformations are linear mappings and are widely used in applications.

Stretching. Different stretch ratios along separate axis in $\mathbb{R}^m$ is described by the linear mapping, $s: \mathbb{R}^m \rightarrow \mathbb{R}^m$,

$$s(\boldsymbol{x}) = \begin{bmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \vdots \\ \lambda_m x_m \end{bmatrix}.$$

The matrix associated with stretching is

$$\boldsymbol{S} = [\, s(\boldsymbol{e}_1) \;\; s(\boldsymbol{e}_2) \;\; \dots \;\; s(\boldsymbol{e}_m) \,] = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_m),$$

and has a remarkably simple form known as a *diagonal matrix*.

Projection. A very important transformation is projection of a vector $\boldsymbol{v} \in \mathbb{R}^m$ along the direction of another vector $\boldsymbol{u} \neq \boldsymbol{0}$. It is convenient to define a vector of unit length along the direction of $\boldsymbol{u}$ by

$$\boldsymbol{q} = \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}.$$

The resulting vector is $\boldsymbol{w} = p_q(\boldsymbol{v}) \in \mathbb{R}^m$ has the same number of components as $\boldsymbol{v}$, and is of length $\|\boldsymbol{v}\| \cos\theta$ in the direction of $\boldsymbol{q}$, stated as

$$\boldsymbol{w} = (\|\boldsymbol{v}\| \cos\theta)\, \boldsymbol{q}.$$

The matrix associated with projection is

$$\boldsymbol{P_q} = [\, p_q(\boldsymbol{e}_1) \;\; p_q(\boldsymbol{e}_2) \;\; \dots \;\; p_q(\boldsymbol{e}_m) \,].$$

The jth unit vector $\boldsymbol{e}_j$ is at angle θ_j with respect to $\boldsymbol{q}$,

$$\cos\theta_j = \boldsymbol{q}^T \boldsymbol{e}_j = [\, q_1 \;\; q_2 \;\; \dots \;\; q_m \,] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = q_j,$$

and the projection of $\boldsymbol{e}_j$ along $\boldsymbol{q}$ is

$$p_q(\boldsymbol{e}_j) = (\|\boldsymbol{e}_j\| \cos\theta_j)\, \boldsymbol{q} = q_j \boldsymbol{q}.$$

Gathering projections of all the unit vectors within the identity matrix gives

$$\boldsymbol{P_q} = [\, q_1 \boldsymbol{q} \;\; q_2 \boldsymbol{q} \;\; \dots \;\; q_m \boldsymbol{q} \,].$$
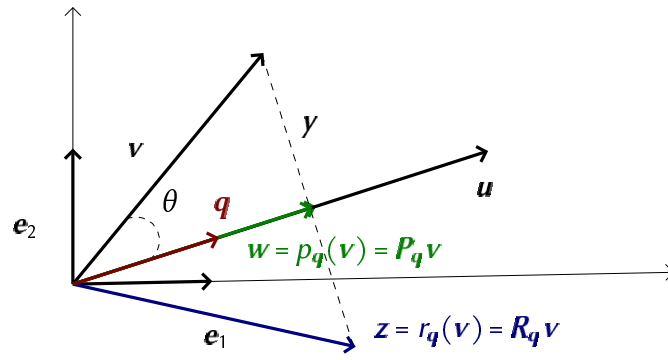
Note that $\boldsymbol{P_q}$ contains m column vectors that are all scalings of the $\boldsymbol{q}$ vector, by coefficients q_1, $q_2, \dots, q_m$ that are the components of $\boldsymbol{q}$ itself. Since scaling is a linear combination, the above m linear combinations can be expressed as a matrix-matrix product

$$\boldsymbol{P_q} = \boldsymbol{q}\, [\, q_1 \;\; q_2 \;\; \dots \;\; q_m \,],$$

leading to the remarkably simple expression

$$\boldsymbol{P_q} = \boldsymbol{q} \boldsymbol{q}^T \in \mathbb{R}^{m \times m}.$$

Rotation in $\mathbb{R}^3$. The axis of the above two dimensinal rotation is a third direction perpendicular to the $x_1 x_2$-plane. A vector $v \in \mathbb{R}^3$ would not change its third coordinate under such a transformation $r_{\theta,3}(v)$ hence the associated rotation matrix is readily obtained as

$$R_\theta = [\ r_{\theta,3}(e_1)\ \ r_{\theta,3}(e_2)\ \ r_{\theta,3}(e_3)\ ] = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3.2. Matrix-matrix product

From two functions $f: A \to B$ and $g: B \to C$, a composite function, $h = g \circ f$, $h: A \to C$ is defined by

$$h(x) = g(f(x)).$$

Consider linear mappings between Euclidean spaces $f: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^m \to \mathbb{R}^p$. Recall that linear mappings are expressed as matrix vector multiplications

$$f(x) = Ax, g(y) = By, A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times m}.$$

The composite function $h = g \circ f$ is $h: \mathbb{R}^n \to \mathbb{R}^p$, defined by

$$h(x) = g(f(x)) = g(Ax) = BAx.$$

Note that the intemediate vector $u = Ax$ is subsequently multiplied by the matrix B. The composite function h is itself a linear mapping

$$h(ax + by) = BA(ax + by) = B(aAx + bAy) = aBAx + bBAy = ah(x) + bh(y),$$

so it also can be expressed a matrix-vector multiplication

$$h(x) = Cx = BAx. \tag{10}$$

Using the above, C is defined as the product of matrix B with matrix A

$$C = BA.$$

The columns of C can be determined from those of A by considering the action of h on the the column vectors of the identity matrix $I = [\ e_1\ \ e_2\ \ \dots\ \ e_n\ ] \in \mathbb{R}^{n \times n}$. First, note that

$$Ae_j = [\ a_1\ \ a_2\ \ \dots\ \ a_n\ ] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} = a_1, \dots, Ae_j = [\ a_1\ \ a_2\ \ \dots\ \ a_n\ ] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_j, Ae_n = [\ a_1\ \ a_2\ \ \dots\ \ a_n\ ] \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} = a_n. \tag{11}$$

The above can be repeated for the matrix $C = [\ c_1\ \ c_2\ \ \dots\ \ c_n\ ]$ giving

$$h(e_1) = Ce_1 = c_1, \dots, h(e_j) = Ce_j = c_j, \dots, h(e_n) = Ce_n = c_n. \tag{12}$$

Combining the above equations leads to $c_j = Ba_j$, or

$$C = [\ c_1\ \ c_2\ \ \dots\ \ c_n\ ] = B[\ a_1\ \ a_2\ \ \dots\ \ a_n\ ].$$

From the above the matrix-matrix product $C = BA$ is seen to simply be a grouping of all the products of B with the column vectors of A,

$$C = [\, c_1 \;\; c_2 \;\; \dots \;\; c_n \,] = [\, B\, a_1 \;\; B a_2 \;\; \dots \;\; B a_n \,]$$

The above results can readily be verified computationally.

```
∴ a1=[1; 2]; a2=[3; 4]; A=[a1 a2]
```

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \tag{13}$$

```
∴ b1=[-1; 1; 3]; b2=[2; -2; 3]; B=[b1 b2]
```

$$\begin{bmatrix} -1 & 2 \\ 1 & -2 \\ 3 & 3 \end{bmatrix} \tag{14}$$

```
∴ C=B*A
```

$$\begin{bmatrix} 3 & 5 \\ -3 & -5 \\ 9 & 21 \end{bmatrix} \tag{15}$$

```
∴ c1=B*a1; c2=B*a2; [c1 c2]
```

$$\begin{bmatrix} 3 & 5 \\ -3 & -5 \\ 9 & 21 \end{bmatrix} \tag{16}$$

```
∴
```

3.3. Properties of the matrix-matrix product

Matrix-matrix products have been seen to arise from composition of linear mappings, and their properties arise from such compositions. Whereas matrix addition is direct consequence of component-by-component operations, matrix products exhibit some particularities.

Non-commutativity. Consider $g(v) = Bv$ to be rotation of vectors in $\mathbb{R}^2$ by angle $\theta = \pi/4$, and $f(v) = Av$ to be reflection across the e_1 direction. The associated matrices are

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Rotation followed by reflection of the vector e_1 is

$$(f \circ g)(e_1) = f(g(e_1)) = ABe_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

whereas reflection followed by rotation is

$$(g \circ f)(e_1) = g(f(e_1)) = BAe_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

a different result, highlighting that the order of applying linear transformations is impor-
tant. Indeed, from

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}, BA = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

it is seen that $AB \neq BA$, and in general matrix multiplication is not commutative.

Associativity. Consider now three linear mappings $f(x) = Ax$, $g(x) = Bx$, $h(x) = Cx$, and
compare the results of $(f \circ g) \circ h$ to $f \circ (g \circ h)$

$$((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = (AB)(Cx)$$

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = A(BCx).$$

Since $(f \circ g)(h(x)) = f(g(h(x))) = (f \circ (g \circ h))(x)$, it results that the above two expres-
sions should also be equal for arbitrary x,

$$(AB)C = A(BC),$$

and matrix multiplication is said to be associative.

Product transposition. The transpose of a linear combination is a straightforward reorga-
nization of vector components as rows instead of columns,

$$(Ax)^T = (x_1 a_1 + \cdots + x_n a_n)^T = x_1 a_1^T + \cdots + x_n a_n^T.$$

The above can be expressed as a matrix multiplication

$$(Ax)^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} = x^T A^T.$$

Extending the above to multiple linear combinations gives

$$(AX)^T = ([Ax_1 \ Ax_2 \ \cdots \ Ax_p])^T = \begin{bmatrix} (Ax_1)^T \\ (Ax_2)^T \\ \vdots \\ (Ax_p)^T \end{bmatrix} = \begin{bmatrix} x_1^T A^T \\ x_2^T A^T \\ \vdots \\ x_p^T A^T \end{bmatrix} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_p^T \end{bmatrix} A^T = X^T A^T,$$

the transpose of a product is the product of the transposes.

3.4. Block matrix operations

In expression such as

$$A^T = ([a_1 \ a_2 \ \cdots \ a_n])^T = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix},$$

the column vectors $a_1, a_2, \ldots, a_n$ can be interpreted as blocks of size $m \times 1$ within the matrix A of size $m \times n$. A pervasive task within linear algebra application to large data sets is to break down the problem into smaller parts. Consider that matrix M of size $m \times n$ can be broken up into blocks

$$M = \begin{bmatrix} U & V \\ X & Y \end{bmatrix}.$$

The dimensions of the blocks have to be compatible, and if U has size $p \times q$ and Y has size $r \times s$, it must hold that

$$m = p + r, n = q + s,$$

and matrix X has size $r \times q$, V has size $p \times s$.

Block transposition. The transpose of M is

$$M^T = \begin{bmatrix} U^T & X^T \\ V^T & Y^T \end{bmatrix}.$$

Block addition. Assuming compatibility of block dimensions, addition is carried out block-by-block.

$$M + N = \begin{bmatrix} U & V \\ X & Y \end{bmatrix} + \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \begin{bmatrix} U + P & V + Q \\ X + R & Y + S \end{bmatrix}.$$

Block multiplication. The "rows-over-columns" rule carries over to blocks

$$MN = \begin{bmatrix} U & V \\ X & Y \end{bmatrix} \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \begin{bmatrix} UP + VR & UQ + VS \\ XP + YR & XQ + YS \end{bmatrix},$$

where non-commutativity of matrix multiplication has to be respected.