Chapter 1
Linear Combinations

Vectors and Matrices

1. Quantities

1.1. Numbers

Most scientific disciplines introduce an idea of the amount of some entity or property of interest. Furthermore, the amount is usually combined with the concept of a *number*, an abstraction of the observation that the two sets $A = \{\text{Mary, Jane, Tom}\}$ and $B = \{\text{apple, plum, cherry}\}$ seem quite different, but we can match one distinct person to one distinct fruit as in $\{\text{Mary}\longrightarrow\text{plum, Jane}\longrightarrow\text{apple, Tom}\longrightarrow\text{cherry}\}$. In contrast we cannot do the same matching of distinct persons to a distinct color from the set {red, green}, and one of the colors must be shared between two persons. Formal definition of the concept of a number from the above observations is surprisingly difficult since it would be self-referential due to the apperance of the numbers "one" and "two". Leaving this aside, the key concept is that of *quantity* of some property of interest that is expressed through a number. Several types of numbers have been introduced in mathematics to express different types of quantities, and the following will be used throughout this text:

- $\mathbb{N}$. The set of natural numbers, $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, infinite and countable, $\mathbb{N}_+ = \{1, 2, 3, \ldots\}$;
- $\mathbb{Z}$. The set of integers, $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$, infinite and countable;
- $\mathbf{Q}$. The set of rational numbers $\mathbb{Q} = \{p/q, p \in \mathbb{Z}, q \in \mathbb{N}_+\}$, infinite and countable;
- $\mathbb{R}$. The set of real numbers, infinite, not countable, can be ordered;
- $\mathbb{C}$. The set of complex numbers, $\mathbb{C} = \{x + iy, x, y \in \mathbb{R}\}$, infinite, not countable, cannot be ordered.

A computer has a finite amount of memory, hence cannot represent all numbers, but rather subsets of the above sets. Furthermore, computers internally use binary numbers composed of binary digits, or bits. Many computer number types are defined for specific purposes, and are often encountered in applications such as image representation or digital data acquisition. Here are the main types.

Subsets of $\mathbb{N}$. The number types `uint8`, `uint16`, `uint32`, `uint64` represent subsets of the natural numbers (unsigned integers) using 8, 16, 32, 64 bits respectively. An unsigned integer with b bits can store a natural number in the range from 0 to $2^b - 1$. Two arbitrary natural numbers, written as $\forall i, j \in \mathbb{N}$ can be added and will give another natural number, $k = i + j \in \mathbb{N}$. In contrast, addition of computer unsigned integers is only defined within the specific range 0 to $2^b - 1$.

```
octave] i=uint8(15); j=uint8(10); k=i+j
```

```
k = 25
```

```
octave] i=uint8(150); j=uint8(200); k=i+j
```

```
k = 255
```

```
octave] k=i-j
```

```
k = 0
```

```
octave]
```

Subsets of $\mathbb{Z}$. The number types `int8`, `int16`, `int32`, `int64` represent subsets of the integers. One bit is used to store the sign of the number, so the subset of $\mathbb{Z}$ that can be represented is from $1 - 2^{b-1}$ to $2^{b-1} - 1$

```
octave] i=int8(100); j=int8(101); k=i+j
```

```
k = 127
```

```
octave] k=i-j
```

```
k = -1
```

```
octave]
```

Subsets of $\mathbb{Q}, \mathbb{R}, \mathbb{C}$. Computers approximate the real numbers through the set $\mathbb{F}$ of *floating point numbers*. Floating point numbers that use $b = 32$ bits are known as *single precision*, while those that use $b = 64$ are *double precision*. A floating point number $x \in \mathbb{F}$ is stored internally as $x = \pm.B_1B_2\ldots B_m \times 2^{\pm b_1b_2\ldots b_e}$ where B_i, $i = 1,\ldots,m$ are bits within the *mantissa* of length m, and b_j, $j = 1,\ldots,e$ are bits within the *exponent*, along with signs $\pm$ for each. The default number type is usually double precision, more concisely referred to `double`. Common constants such as e, π are predefined as `double`, can be truncated to `single`, and the number of displayed decimal digits is controlled by `format`. The function `disp(x)` displays its argument `x`.

```
octave] format long; disp([e pi])
```

```
   2.718281828459045   3.141592653589793
```

```
octave] disp([single(e) single(pi)])
```

```
   2.7182817   3.1415927
```

```
octave]
```

The approximation of the reals $\mathbb{R}$ by the floats $\mathbb{F}$ is characterized by: `realmax`, the largest float, `realmin` the smallest positive float, and `eps` known as *machine epsilon*. Machine epsilon highlights the differences between floating point and real numbers since it is defined as the largest number $\epsilon \in \mathbb{F}$ that satisfies $1 + \epsilon = 1$. If $\varepsilon \in \mathbb{R}$ of course $1 + \varepsilon = 1$ implies $\varepsilon = 0$, but floating points exhibit "granularity", in the sense that over a unit interval there are small steps that are indistinguishable from zero due to the finite number of bits available for a float. Machine epsilon is small, and floating point errors can usually be kept under control. Keep in mind that perfect accuracy is a mathematical abstraction, not encountered in nature. In fields as sociology or psychology 3 digits of accuracy are excellent, in mechanical engineering this might increase to 6 digits, or in electronic engineering to 8 digits. The most precisely known physical constant is the Rydberg constant known to 12 digits. The granularity of double precision expressed by machine epsilon is sufficient to represent natural phenomena.

```
octave] format short; disp([realmin realmax eps 1+eps])
```

```
   2.2251e-308   1.7977e+308    2.2204e-16    1.0000e+00
```

```
octave]
```

Within the reals certain operations are undefined such as $1/0$. Special float constants are defined to handle such situations: `Inf` is a float meant to represent infinity, and `NaN` ("not a number") is meant to represent an undefinable result of an arithmetic operation.

```
octave] warning("off"); disp([Inf 1/0 2*realmax NaN Inf-Inf Inf/Inf])
```

```
   Inf   Inf   Inf   NaN   NaN   NaN
```

```
octave]
```

Complex numbers $z \in \mathbb{C}$ are specified by two reals, in Cartesian form as $z = x + iy$, $x, y \in \mathbb{R}$ or in polar form as $z = \rho e^{i\theta}$, $\rho, \theta \in \mathbb{R}$, $\rho \geqslant 0$. The computer type complex is similarly defined from two floats and the additional constant I is defined to represent $\sqrt{-1} = i = e^{i\pi/2}$. Functions are available to obtain the real and imaginary parts within the Cartesian form, or the absolute value and argument of the polar form.

```
octave] z1=complex(1,1); z2=complex(1,-1); disp([z1+z2 z1/z2])
```

```
   2 + 0i   0 + 1i
```

```
octave] disp([real(z1) real(z2) real(z1+z2) real(z1/z2)])
```

```
   1   1   2   0
```

```
octave] disp([imag(z1) imag(z2) imag(z1+z2) imag(z1/z2)])
   1  -1   0   1
octave] disp([abs(z1) abs(z2) abs(z1+z2) abs(z1/z2)])
   1.4142   1.4142   2.0000   1.0000
octave] disp([arg(z1) arg(z2) arg(z1+z2) arg(z1/z2)])
   0.78540  -0.78540   0.00000   1.57080
octave] I-sqrt(-1)
ans = 0
octave]
```

Care should be exercised about the cummulative effect of many floating point errors. For instance, in an "irrational" numerical investigation of Zeno's paradox, one might want to compare the distance S_N traversed by step sizes that are scaled by $1/\pi$ starting from one to T_N, traversed by step sizes scaled by π starting from π^{-N}

$$S_N = 1 + \frac{1}{\pi} + \frac{1}{\pi^2} + \cdots + \frac{1}{\pi^N}, \; T_N = \frac{1}{\pi^N} + \frac{1}{\pi^{N-1}} + \cdots + 1.$$

In the reals the above two expressions are equal, $S_N = T_N$, but this is not verfied for all N when using floating point numbers. Lists of the values π^j, for the two orderings $j = 0, \ldots, N$, and $j = N, \ldots, 0$, can be generated and summed.

```
octave] N=10; S=pi.^(0:-1:-N); T=pi.^(-N:1:0); sum(S)==sum(T)
ans = 1
octave] N=15; S=pi.^(0:-1:-N); T=pi.^(-N:1:0); sum(S)==sum(T)
ans = 0
octave]
```

In the above numerical experiment a==b expresses an equality relationship which might evaluate as true denoted by 1, or false denoted by 0.

```
octave] disp([1==1 1==2])
  1  0
octave]
```

The above was called an "irrational" investigation since in Zeno's original paradox the scaling factor was 2 rather than π, and due to the binary representation used by floats equality always holds.

```
octave] N=30; S=2.^(0:-1:-N); T=2.^(-N:1:0); sum(S)==sum(T)
ans = 1
octave]
```

1.2. Quantities described by a single number

The above numbers and their computer approximations are sufficient to describe many quantities encountered in applications. Typical examples include:

- the position $x \in \mathbb{R}$ of a point on the unit line segment $[0, 1]$, approximated by the floating point number $\tilde{x} \in \mathbb{F}$, to within machine epsilon precision, $|x - \tilde{x}| \leqslant \epsilon$;
- the measure of resistance to change of the rate of motion known as *mass*, $m \in \mathbb{R}$, $m > 0$;
- the population of a large community expressed as a float $p \in \mathbb{F}$, even though for a community of individuals the population is a natural number, as in "the population of the United States is $p = 328.2\text{E}6$, i.e., 328.2 million".

In most disciplines, there is a particular interest in comparison of two quantities, and to facilitate such comparison a common reference is used known as a *standard unit*. For measurement of a length L, the meter $\ell = 1\,\mathrm{m}$ is a standard unit, as in the statement $L = 10\,\mathrm{m}$, that states that L is obtained by taking the standard unit ten times, $L = 10\,\ell$. The rules for carrying out such comparisons are part of the definition of real and rational numbers. These rules are formalized in the mathematical definition of a *field* $(F, +, \times)$ presented in the next chapter. Quantities that obey such rules, i.e., belong to a field, can be used in changes of scale and are called *scalars*. Not all numbers are scalars in this sense. For instance, the integers would not allow a scaling of 1:2 (halving the scale) even though 1,2 are integers.

1.3. Quantities described by multiple numbers

Other quantities require more than a single number. The distribution of population in the year 2000 among the alphabetically-ordered South American countries (Argentina, Bolivia,..,Venezuela) requires 12 numbers. These are placed together in a list known in mathematics as a *tuple*, in this case a 12-tuple $P = (p_1, p_2, \ldots, p_{12})$, with p_1 the population of Argentina, p_2 that of Bolivia, and so on. An analogous 12-tuple can be formed from the South American populations in the year 2020, say $Q = (q_1, q_2, \ldots, q_{12})$. Note that it is difficult to ascribe meaning to apparently plausible expressions such as $P + Q$ since, for instance, some people in the 2000 population are also in the 2020 population, and would be counted twice.

2. Vectors

2.1. Vector spaces

In contrast to the population 12-tuple example above, combining multiple numbers is well defined in operations such as specifying a position within a three-dimensional Cartesian grid, or determining the resultant of two forces in space. Both of these lead to the consideration of 3-tuples or *triples* such as the force (f_1, f_2, f_3). When combined with another force (g_1, g_2, g_3) the resultant is $(f_1 + g_1, f_2 + g_2, f_3 + g_3)$. If the force (f_1, f_2, f_3) is amplified by the scalar α and the force (g_1, g_2, g_3) is similarly scaled by β, the resultant becomes

$$\alpha(f_1, f_2, f_3) + \beta(g_1, g_2, g_3) = (\alpha f_1, \alpha f_2, \alpha f_3) + (\beta g_1, \beta g_2, \beta g_3) = (\alpha f_1 + \beta g_1, \alpha f_2 + \beta g_2, \alpha f_3 + \beta g_3).$$

It is useful to distinguish tuples for which scaling and addition is well defined from simple lists of numbers. In fact, since the essential difference is the behavior with respect to scaling and addition, the focus should be on these operations rather than the elements of the tuple.

The above observations underlie the definition of a *vector space* $\mathscr{V}$ by a set V whose elements satisfy certain scaling and addition properties, denoted all together by the 4-tuple $\mathscr{V} = (V, S, +, \cdot)$. The first element of the 4-tuple is a set whose elements are called *vectors*. The second element is a set of scalars, and the third is the vector addition operation. The last is the scaling operation, seen as multiplication of a vector by a scalar. The vector addition and scaling operations must satisfy rules suggested by positions or forces in three-dimensional space, which are listed in Table ?. In particular, a vector space requires definition of two distinguished elements: the zero vector $\mathbf{0} \in V$, and the identity scalar element $1 \in S$.

Addition rules for	$\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$
$\boldsymbol{a} + \boldsymbol{b} \in V$	Closure
$\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}$	Associativity
$\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$	Commutativity
$\mathbf{0} + \boldsymbol{a} = \boldsymbol{a}$	Zero vector
$\boldsymbol{a} + (-\boldsymbol{a}) = \mathbf{0}$	Additive inverse
Scaling rules for	$\forall \boldsymbol{a}, \boldsymbol{b} \in V, \forall x, y \in S$
$x\boldsymbol{a} \in V$	Closure
$x(\boldsymbol{a} + \boldsymbol{b}) = x\boldsymbol{a} + x\boldsymbol{b}$	Distributivity
$(x + y)\boldsymbol{a} = x\boldsymbol{a} + y\boldsymbol{a}$	Distributivity
$x(y\boldsymbol{a}) = (xy)\boldsymbol{a}$	Composition
$1\boldsymbol{a} = \boldsymbol{a}$	Scalar identity

Table 1.1. Vector space $\mathscr{V} = (V, S, +, \cdot)$ properties for arbitrary $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$

The definition of a vector space reflects everyday experience with vectors in Euclidean geometry, and it is common to refer to such vectors by descriptions in a Cartesian coordinate system. For example, a position vector $\boldsymbol{r}$ within the plane can be referred through the pair of coordinates (x, y). This intuitive understanding can be made precise through the definition of a vector space $\mathscr{R}_2 = (\mathbb{R}^2, \mathbb{R}, +, \cdot)$, called the real 2-space. Vectors within $\mathscr{R}_2$ are elements of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$, meaning that a vector is specified through two real numbers, $\boldsymbol{r} \leftrightarrow (x, y)$. Addition of two vectors, $\boldsymbol{q} \leftrightarrow (s, t)$, $\boldsymbol{r} \leftrightarrow (x, y)$ is defined by addition of coordinates $\boldsymbol{q} + \boldsymbol{r} = (s + x, t + v)$. Scaling $\boldsymbol{r} \leftrightarrow (x, y)$ by scalar a is defined by $a\boldsymbol{r} \leftrightarrow (ax, ay)$. Similarly, consideration of position vectors in three-dimensional space leads to the definition of the $\mathscr{R}_3 = (\mathbb{R}^3, \mathbb{R}, +, \cdot)$, or more generally a real m-space $\mathscr{R}_m = (\mathbb{R}^m, \mathbb{R}, +, \cdot)$, $m \in \mathbb{N}$, $m > 0$.

Note however that there is no mention of coordinates in the definition of a vector space as can be seen from the list of properties in Table 1.1. The intent of such a definition is to highlight that besides position vectors, many other mathematical objects follow the same rules. As an example, consider the set of all continuous functions $C(\mathbb{R}) = \{f \mid f : \mathbb{R} \longrightarrow \mathbb{R}\}$, with function addition defined by the sum at each argument t, $(f+g)(t) = f(t) + g(t)$, and scaling by $a \in \mathbb{R}$ defined as $(af)(t) = af(t)$. Read this as: "given two continuous functions f and g, the function $f+g$ is defined by stating that its value for argument x is the sum of the two real numbers $f(t)$ and $g(t)$". Similarly: "given a continuous function f, the function af is defined by stating that its value for argument t is the product of the real numbers a and $f(t)$". Under such definitions $\mathscr{C}^0 = (C(\mathbb{R}), \mathbb{R}, +, \cdot)$ is a vector space, but quite different from $\mathscr{R}_m$. Nonetheless, the fact that both $\mathscr{C}^0$ and $\mathscr{R}_m$ are vector spaces can be used to obtain insight into the behavior of continuous functions from Euclidean vectors, and vice versa.

2.2. Real vector space $\mathscr{R}_m$

Column vectors. Since the real spaces $\mathscr{R}_m = (\mathbb{R}^m, \mathbb{R}, +, \cdot)$ play such an important role in themselves and as a guide to other vector spaces, familiarity with vector operations in $\mathscr{R}_m$ is necessary to fully appreciate the utility of linear algebra to a wide range of applications. Following the usage in geometry and physics, the m real numbers that specify a vector $\boldsymbol{u} \in \mathbb{R}^m$ are called the *components* of $\boldsymbol{u}$. The one-to-one correspondence between a vector and its components $\boldsymbol{u} \leftrightarrow (u_1, \ldots, u_m)$, is by convention taken to define an equality relationship,

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}, \tag{1.1}$$

with the components arranged vertically and enclosed in square brackets. Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^m$, and a scalar $a \in \mathbb{R}$, vector addition and scaling are defined in $\mathscr{R}_m$ by real number addition and multiplication of components

$$\boldsymbol{u} + \boldsymbol{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_m + v_m \end{bmatrix}, a\boldsymbol{u} = a \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \begin{bmatrix} a u_1 \\ \vdots \\ a u_m \end{bmatrix}. \tag{1.2}$$

The vector space $\mathscr{R}_m$ is defined using the real numbers as the set of scalars, and constructing vectors by grouping together m scalars, but this approach can be extended to any set of scalars S, leading to the definition of the vector spaces $\mathscr{S}_n = (S^n, S, +, \cdot)$. These will often be referred to as *n-vector space of scalars*, signifying that the set of vectors is $V = S^n$.

To aid in visual recognition of vectors, the following notation conventions are introduced:

- vectors are denoted by lower-case bold Latin letters: $\boldsymbol{u}, \boldsymbol{v}$;
- scalars are denoted by normal face Latin or Greek letters: a, b, α, β;
- the components of a vector are denoted by the corresponding normal face with subscripts as in equation (1.1);
- related sets of vectors are denoted by indexed bold Latin letters: $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$.

In Octave, successive components placed vertically are separated by a semicolon.

```
octave] [1; 2; -1; 2]
```

```
ans =

   1
   2
  -1
   2
```

```
octave]
```

The equal sign in mathematics signifies a particular equivalence relationship. In computer systems such as Octave the equal sign has the different meaning of *assignment*, that is defining the label on the left side of the equal sign to be the expression on the right side. Subsequent invocation of the label returns the assigned object. Components

of a vector are obtained by enclosing the index in parantheses.

```
octave] u=[1; 2; -1; 2]; u
```

```
u =

   1
   2
  -1
   2
```

```
octave] u(3)
```

```
ans = -1
```

```
octave]
```

Row vectors. Instead of the vertical placement or components into one *column*, the components of could have been placed horizontally in one *row* $[\ u_1\ \ldots\ u_m\ ]$, that contains the same data, differently organized. By convention vertical placement of vector components is the preferred organization, and $\boldsymbol{u}$ shall denote a *column vector* henceforth. A transpose operation denoted by a T superscript is introduced to relate the two representations

$$\boldsymbol{u}^T = [\ u_1\ \ldots\ u_m\ ],$$

and $\boldsymbol{u}^T$ is the notation used to denote a *row vector*. In Octave, horizontal placement of successive components in a row is denoted by a space.

```
octave] uT=transpose(u)
```

```
uT =

   1   2  -1   2
```

```
octave] [1 2 -1 2]
```

```
ans =

   1   2  -1   2
```

```
octave] uT(4)
```

```
ans =  2
```

```
octave]
```

Compatible vectors. Addition of real vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^m$ defines another vector $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v} \in \mathbb{R}^m$. The components of $\boldsymbol{w}$ are the sums of the corresponding components of $\boldsymbol{u}$ and $\boldsymbol{v}$, $w_i = u_i + v_i$, for $i = 1, 2, \ldots, m$. Addition of vectors with different number of components is not defined, and attempting to add such vectors produces an error. Such vectors with different number of components are called *incompatible*, while vectors with the same number of components are said to be *compatible*. Scaling of $\boldsymbol{u}$ by a defines a vector $\boldsymbol{z} = a\boldsymbol{u}$, whose components are $z_i = a u_i$, for $i = 1, 2, \ldots, m$. Vector addition and scaling in Octave are defined using the + and $*$ operators.

```
octave] uT=[1 0 1 2]; vT=[2 1 3 -1]; wT=uT+vT; disp(wT)
```

```
   3   1   4   1
```

```
octave] rT=[1 2]; uT+rT
```

```
operator +: nonconformant arguments (op1 is 1x4, op2 is 1x2)
```

```
octave] a=3; zT=a*uT; disp(zT)
   3   6  -3   6
octave]
```

2.3. Working with vectors

Ranges. The vectors used in applications usually have a large number of components, $m \gg 1$, and it is important to become proficient in their manipulation. Previous examples defined vectors by explicit listing of their m components. This is impractical for large m, and support is provided for automated generation for often-encountered situations. First, observe that Table 1.1 mentions one distinguished vector, the zero element that is a member of any vector space $\mathbf{0} \in V$. The zero vector of a real vector space $\mathscr{R}_m$ is a column vector with m components, all of which are zero, and a mathematical convention for specifying this vector is $\mathbf{0}^T = [\ 0\ \ 0\ \ \ldots\ \ 0\ ] \in \mathbb{R}^m$. This notation specifies that transpose of the zero vector is the row vector with m zero components, also written through explicit indexing of each component as $\mathbf{0}_i = 0$, for $i = 1, \ldots, m$. Keep in mind that the zero vector $\mathbf{0}$ and the zero scalar 0 are different mathematical objects. The ellipsis symbol in the mathematical notation is transcribed in Octave by the notion of a range, with `1:m` denoting all the integers starting from 1 to m, organized as a row vector. The notation is extended to allow for strides different from one, and the mathematical ellipsis $i = m, m-1, \ldots, 1$ is denoted as `m:-1:1`. In general `r:s:t` denotes the set of numbers $\{r, r+s, \ldots, r+ns\}$ with $r + ns \leqslant t$, and r, s, t real numbers and n a natural number, $r, s, t \in \mathbb{R}$, $n \in \mathbb{N}$. If there is no natural number n such that $r + ns \leqslant t$, an empty vector with no components is returned.

```
octave] m=4; disp(1:m)
   1   2   3   4
octave] disp(m:-1:2)
   4   3   2
octave] r=0; s=0.2; t=1; disp(r:s:t)
    0.00000    0.20000    0.40000    0.60000    0.80000    1.00000
octave] r=0; s=0.3; t=1; disp(r:s:t)
    0.00000    0.30000    0.60000    0.90000
octave] r=0; s= -0.2; t=1; disp(r:s:t)
[](1x0)
octave]
```

An efficient, expressive feature of many software systems including Octave is to use ranges as indices to a vector, as shown below for the definition of $\mathbf{0} \in \mathbb{R}^4$. Note that the index range `i` is organized as a row, and a transpose operation must be applied to obtain `z` as a column vector.

```
octave] m=4; i=1:m; z(i)=i.^i; z=transpose(z); disp(z)
     1
     4
    27
   256
octave] i
i =

   1   2   3   4
```

```
octave] disp(transpose(z))
   0   0   0   0
octave]
```

Visualization. A component-by-component display of a vector becomes increasingly unwieldy as the number of components m becomes large. For example, the numbers below seem an inefficient way to describe the sine function.

```
octave]  t=0:0.1:1.5; disp(sin(t))
  Columns 1 through 8:

    0.00000   0.09983   0.19867   0.29552   0.38942   0.47943   0.56464   0.64422

  Columns 9 through 16:

    0.71736   0.78333   0.84147   0.89121   0.93204   0.96356   0.98545   0.99749
octave]
```

Indeed, such a piece-by-piece approach is not the way humans organize large amounts of information, preferring to conceptualize the data as some other entity: an image, a sound excerpt, a smell, a taste, a touch, a sense of balance, or relative position. All seven of these human senses will be shown to allow representation by linear algebra concepts, including representation by vectors.

As a first example consider *visualization*, the process of transforming data into a sight perception. A familiar example is constructing a plot of the graph of a function. Recall that in mathematics the *graph of a function* $f: X \longrightarrow Y$ relating elements of the *domain* X to those in *codomain* Y is a set of ordered pairs $G_f = \{(x, y) \mid y = f(x), x \in X\}$. For a commonly encountered function such as $\sin: [0, 2\pi) \longrightarrow [-1, 1]$, the graph $G_{\sin} = \{(x, \sin(x)) \mid x \in [0, 2\pi)\}$ contains an uncountably infinite number of elements, and obviously cannot be explicitly listed. The sine function is continuous, meaning that no matter how small an open interval (c, d) within the function codomain $[-1, 1]$ one considers, there exists an interval (a, b) in the function domain $[0, 2\pi]$ whose image by the sine function is contained in (c, d). In mathematical "$\delta - \varepsilon$" notation this is stated as: $\forall \varepsilon > 0, \exists \delta_\varepsilon, \mid x_1 - x_0 \mid < \delta_\varepsilon \Longrightarrow \mid \sin(x_1) - \sin(x_0) \mid < \varepsilon$. This mathematical notation is concise and precise, but perceptive mainly to the professional mathematician. A more intuitive visualization of continuity is obtained by approximating the graph of a function $f: X \longrightarrow Y$ by a finite set of samples, $G_f^m = \{(x_i, y_i) \mid x_i \in X, y_i = f(x_i), i = 1, \ldots, m, m \in \mathbb{N}\}$. Strictly speaking, the sampled graph G_f^m would indicate jumps interpretable as discontinuities, but when plotting the points human sight perception conveys a sense of continuity for large sample sizes, $m \gg 1$. For the sine function example, consider sampling the domain $[0, 2\pi)$ with a step size $h = 2\pi / m$, $m \gg 1$. To obtain a visual representation of the sampled sine function the Octave `plot` function can be used to produce a figure that will appear in another window, interactively investigated, and subsequently closed. For large m one cannot visually distinguish the points in the graph sample, though this is apparent for smaller sample sizes. This is shown below by displaying a subrange of the sampled points with stride s. This example also shows the procedure to save a permanent copy of the displayed figure through the Octave `print -deps` command that places the currently displayed plot into an Encapsulated Postscript file. The generated figure file can be linked to a document as shown here in Figure ?, in which both plots render samples of the graph of the sine function, but the one with large m is perceived as being continuous.

```
octave] m=1000; h=2*pi/m; x=(0:m-1)*h;
octave] y=sin(x); plot(x,y);
octave] close;
octave] s=50; i=1:s:m; xs=x(i); ys=y(i);
```

```
octave] plot(x,y,'b',xs,ys,'bo');
octave] print -depsc L01Fig01.eps;
octave] close;
octave]
```

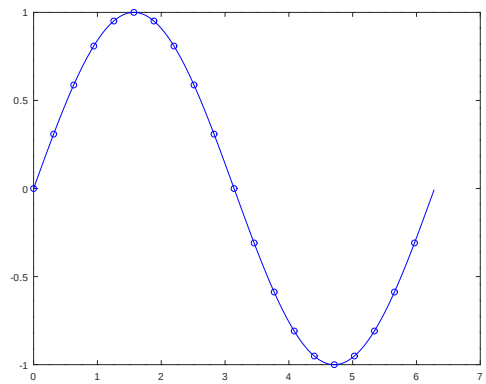

Figure 1.1. Visualization of vectors of sampled function graphs.

3. Matrices

3.1. Forming matrices

The real numbers themselves form the vector space $\mathscr{R}_1 = (\mathbb{R}, \mathbb{R}, +, \cdot)$, as does any field of scalars, $\mathscr{S}_1 = (S, S, +, \cdot)$. Juxtaposition of m real numbers has been seen to define the new vector space $\mathscr{R}_m$. This process of juxtaposition can be continued to form additional mathematical objects. A *matrix* is defined as a juxtaposition of compatible vectors. As an example, consider n vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in V$ within some vector space $\mathscr{V} = (V, S, +, \cdot)$. Form a matrix by placing the vectors into a row,

$$\boldsymbol{A} = \left[\begin{array}{cccc} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_n \end{array}\right]. \tag{1.3}$$

To aid in visual recognition of a matrix, upper-case bold Latin letters will be used to denote matrices. The columns of a matrix will be denoted by the corresponding lower-case bold letter with a subscripted index as in equation (1.3). Note that the number of columns in a matrix can be different from the number of components in each column, as would be the case for matrix $\boldsymbol{A}$ from equation (1.3) when choosing vectors from, say, the real space $\mathscr{R}_m$, $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$.

Vectors were seen to be useful juxtapositions of scalars that could describe quantities a single scalar could not: a position in space, a force in physics, or a sampled function graph. The crucial utility of matrices is their central role in providing a description of new vectors other then their column vectors, and is suggested by experience with Euclidean spaces.

3.2. Identity matrix

Consider first $\mathscr{R}_1$, the vector space of real numbers. A position vector $\boldsymbol{r} \in \mathscr{R}_1$ on the real axis is specified by a single scalar component, $\boldsymbol{r} = [x]$, $x \in \mathbb{R}$. Read this to mean that the position $\boldsymbol{r}$ is obtained by traveling x units from the origin at position vector $\mathbf{0} = [0]$. Look closely at what is meant by "unit" in this context. Since x is a scalar, the mathematical expression $\boldsymbol{r} = \mathbf{0} + x$ has no meaning, as addition of a vector to a scalar has not been defined. Recall that scalars were introduced to capture the concept of scaling of a vector, so in the context of vector spaces they always appear as multiplying some vector. The correct mathematical description is $\boldsymbol{r} = \mathbf{0} + x\,\boldsymbol{e}$, where $\boldsymbol{e}$ is the unit vector $\boldsymbol{e} = [1]$. Taking the components leads to $r_1 = 0_1 + x e_1$, where $r_1, 0_1, e_1$ are the first (and in this case only) components of the $\boldsymbol{r}, \mathbf{0}, \boldsymbol{e}$ vectors. Since $r_1 = x$, $0_1 = 0$, $e_1 = 1$, one obtains the identity $x = 0 + x \cdot 1$.

Now consider $\mathscr{R}_2$, the vector space of positions in the plane. Repeating the above train of thought leads to the identification of two direction vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$

$$\boldsymbol{r} = \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x\,\boldsymbol{e}_1 + y\,\boldsymbol{e}_2,\ \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

```
octave] x=2; y=4; e1=[1; 0]; e2=[0; 1]; r=x*e1+y*e2
```

```
r =

   2
   4
```

```
octave]
```

Continuing the procees to $\mathscr{R}_m$ gives

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = x_1\,\boldsymbol{e}_1 + x_2\,\boldsymbol{e}_2 + \cdots + x_m\,\boldsymbol{e}_m,\ \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},\ \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ldots, \boldsymbol{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

For arbitrary m, the components are now $x_1, x_2, \ldots, x_m$ rather than the alphabetically ordered letters common for $m = 2$ or $m = 3$. It is then consistent with the adopted notation convention to use $\boldsymbol{x} \in \mathscr{R}_m$ to denote the position vector whose components are $(x_1, \ldots, x_m)$. The basic idea is the same as in the previous cases: to obtain a position vector scale direction $\boldsymbol{e}_1$ by x_1, $\boldsymbol{e}_2$ by x_2,..., $\boldsymbol{e}_m$ by x_m, and add the resulting vectors.

Juxtaposition of the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ leads to the formation of a matrix of special utility known as the *identity matrix*

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_m \end{bmatrix}.$$

The identity matrix is an example of a matrix in which the number of column vectors n is equal to the number of components in each column vector $m = n$. Such matrices with equal number of columns and rows are said to be *square*. Due to entrenched practice an exception to the notation convention is made and the identity matrix is denoted by $\boldsymbol{I}$, but its columns are denoted the indexed bold-face of a different lower-case letter, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$. If it becomes necessary to explicitly state the number of columns in $\boldsymbol{I}$, the notation $\boldsymbol{I}_m$ is used to denote the identity matrix with m columns, each with m components.

4. Linear combinations

4.1. Linear combination as a matrix-vector product

The expression $\boldsymbol{x} = x_1\,\boldsymbol{e}_1 + x_2\,\boldsymbol{e}_2 + \cdots + x_m\,\boldsymbol{e}_m$ expresses the idea of scaling vectors within a set and subsequent addition to form a new vector $\boldsymbol{x}$. The matrix $\boldsymbol{I} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_m \end{bmatrix}$ groups these vectors together in a single entity, and the scaling factors are the components of the vector $\boldsymbol{x}$. To bring all these concepts together it is natural to consider the notation

$$\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x},$$

as a generalization of the scalar expression $x = 1 \cdot x$. It is clear what the operation $\boldsymbol{I}\boldsymbol{x}$ should signify: it should capture the vector scaling and subsequent vector addition $x_1\,\boldsymbol{e}_1 + x_2\,\boldsymbol{e}_2 + \cdots + x_m\,\boldsymbol{e}_m$. A specific meaning is now ascribed to $\boldsymbol{I}\boldsymbol{x}$ by identifying two definitions to one another.

Linear combination. Repeateadly stating "vector scaling and subsequent vector addition" is unwieldy, so a special term is introduced for some given set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.

DEFINITION. (LINEAR COMBINATION) . *The linear combination of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in V$ with scalars $x_1, x_2, \ldots, x_n \in S$ in vector space $(V, S, +, \cdot)$ is the vector $\boldsymbol{b} = x_1\,\boldsymbol{a}_1 + x_2\,\boldsymbol{a}_2 + \ldots x_n\,\boldsymbol{a}_n$.*

Matrix-vector product. Similar to the grouping of unit vectors $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m$ into the identity matrix $\boldsymbol{I}$, a more concise way of referring to arbitrary vectors $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ from the same vector space is the matrix $\boldsymbol{A}=[\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \ldots\ \ \boldsymbol{a}_n\ ]$. Combining these observations leads to the definition of a matrix-vector product.

DEFINITION. (MATRIX-VECTOR PRODUCT) . *In the vector space $(V,S,+,\cdot)$, the product of matrix $\boldsymbol{A}=[\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \ldots\ \ \boldsymbol{a}_n\ ]$ composed of columns $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\in V$ with the vector $\boldsymbol{x}\in S_n$ whose components are scalars $x_1,x_2,\ldots,x_n\in S$ is the linear combination $\boldsymbol{b}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\ldots x_n\boldsymbol{a}_n=\boldsymbol{A}\boldsymbol{x}\in V$.*

4.2. Linear algebra problem examples

Linear combinations in E_2. Consider a simple example that leads to a common linear algebra problem: decomposition of forces in the plane along two directions. Suppose a force is given in terms of components along the Cartesian x,y-axes, $\boldsymbol{b}=b_x\boldsymbol{e}_x+b_y\boldsymbol{e}_y$, as expressed by the matrix-vector multiplication $\boldsymbol{b}=\boldsymbol{I}\boldsymbol{b}$. Note that the same force could be obtained by linear combination of other vectors, for instance the normal and tangential components of the force applied on an inclined plane with angle θ, $\boldsymbol{b}=x_t\boldsymbol{e}_t+x_n\boldsymbol{e}_n$, as in Figure ?. This defines an alternate reference system for the problem. The unit vectors along these directions are

$$\boldsymbol{t}=\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix},\boldsymbol{n}=\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix},$$

and can be combined into a matrix $\boldsymbol{A}=[\ \boldsymbol{t}\ \ \boldsymbol{n}\ ]$. The value of the components (x_t,x_n) are the scaling factors and can be combined into a vector $\boldsymbol{x}=[\ x_t\ \ x_n\ ]^T$. The same force must result irrespective of whether its components are given along the Cartesian axes or the inclined plane directions leading to the equality

$$\boldsymbol{I}\boldsymbol{b}=\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}. \tag{1.4}$$

Interpret equation (1.4) to state that the vector $\boldsymbol{b}$ could be obtained either as a linear combination of $\boldsymbol{I}$, $\boldsymbol{b}=\boldsymbol{I}\boldsymbol{b}$, or as a linear combination of the columns of $\boldsymbol{A}$, $\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}$. Of course the simpler description seems to be $\boldsymbol{I}\boldsymbol{b}$ for which the components are already known. But this is only due to an arbitrary choice made by a human observer to define the force in terms of horizontal and vertical components. The problem itself suggests that the tangential and normal components are more relevant; for instance a friction force would be evaluated as a scaling of the normal force. The components in this more natural reference system are not known, but can be determined by solving the vector equality $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, known as a *linear system of equations*. Procedures to carry this out will be studied in more detail later, but Octave provides an instruction for this common problem, the backslash operator, as in `x=A\b`.

```
octave] ex=[1; 0]; ey=[0; 1];
octave] b=[0.2; 0.4]; I=[ex ey]; I*b
ans =

   0.20000
   0.40000
```

```
octave] th=pi/6; c=cos(th); s=sin(th);
octave] tvec=[c; s]; nvec=[-s; c];
octave] A=[tvec nvec];
octave] x=A\b
x =

   0.37321
   0.24641
```

```
octave] [x(1)*tvec x(2)*nvec]
ans =

   0.32321  -0.12321
   0.18660   0.21340

octave]
```

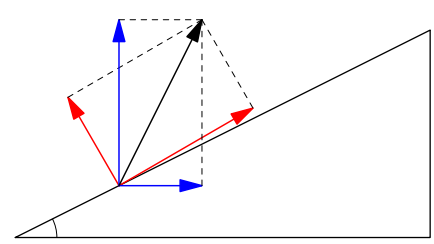

Figure 1.2. Alternative decompositions of force on inclined plane.

Linear combinations in $\mathcal{R}_m$ and $\mathscr{C}^0[0,2\pi)$. Linear combinations in a real space can suggest properties or approximations of more complex objects such as continuous functions. Let $\mathscr{C}^0[0,2\pi) = (C[0,2\pi),\mathbb{R},+,\cdot)$ denote the vector space of continuous functions that are periodic on the interval $[0,2\pi)$, $C[0,\pi) = \{f \mid f:\mathbb{R}\longrightarrow\mathbb{R}, f(t) = f(t+2\pi)\}$. Recall that vector addition is defined by $(f+g)(t) = f(t)+g(t)$, and scaling by $(af)(t) = af(t)$, for $f,g\in C[0,2\pi)$, $a\in\mathbb{R}$. Familiar functions within this vector space are $\sin(kt)$, $\cos(kt)$ with $k\in\mathbb{N}$, and these can be recognized to intrinsically represent periodicity on $[0,2\pi)$, a role analogous to the normal and tangential directions in the inclined plane example.

Define now another periodic function $b(t+2\pi) = b(t)$ by repeating the values $b(t) = t(\pi-t)(2\pi-t)$ from the interval $[0,2\pi)$ on all intervals $[2p\pi, 2(p+1)\pi]$, for $p\in\mathbb{Z}$. The function b is not given in terms of the "naturally" periodic functions $\sin(kt)$, $\cos(kt)$, but could it thus be expressed? This can be stated as seeking a linear combination $b(t) = \sum_{k=1}^{\infty} x_k \sin(kt)$, as studied in Fourier analysis. The coefficients x_k could be determined from an analytical formula involving calculus operations $x_k = \frac{1}{\pi}\int_0^{2\pi} b(t)\sin(kt)\,\mathrm{d}t$, but we'll seek an approximation using a linear combination of n terms

$$b(t) \cong \sum_{k=1}^{n} x_k \sin(kt), A(t) = [\ \sin(t)\ \ \sin(2t)\ \ \ldots\ \ \sin(nt)\ ], A:\mathbb{R}\longrightarrow\mathbb{R}^n.$$

Organize this as a matrix vector product $b(\boldsymbol{t}) \cong A(\boldsymbol{t})\,\boldsymbol{x}$, with

$$A(\boldsymbol{t}) = [\ \sin(\boldsymbol{t})\ \ \sin(2\boldsymbol{t})\ \ \ldots\ \ \sin(n\boldsymbol{t})\ ], \boldsymbol{x} = [\ x_1\ \ x_2\ \ \ldots\ \ x_n\ ]^T \in \mathbb{R}^n.$$

The idea is to sample the column vectors of $A(\boldsymbol{t})$ at the components of the vector $\boldsymbol{t} = [\ t_1\ \ t_2\ \ \ldots\ \ t_m\ ]^T \in \mathbb{R}^m$, $t_j = (j-1)h$, $j = 1,2,\ldots,m$, $h = \pi/m$. Let $\boldsymbol{b} = b(\boldsymbol{t})$, and $\boldsymbol{A} = A(\boldsymbol{t})$, denote the so-sampled b, A functions leading to the definition of a vector $\boldsymbol{b}\in\mathbb{R}^m$ and a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$. There are n coefficients available to scale the column vectors of $\boldsymbol{A}$, and $\boldsymbol{b}$ has m components. For $m > n$ it is generally not possible to find $\boldsymbol{x}$ such that $\boldsymbol{Ax}$ would exactly equal $\boldsymbol{b}$, but as seen later the condition to be as close as possible to $\boldsymbol{b}$ leads to a well defined solution procedure. This is known as a least squares problem and is automatically applied in the Octave `x=A\b` instruction when the matrix `A` is not square. As seen in the following numerical experiment and Figure ?, the approximation is excellent even though the information conveyed by $m = 1000$ samples of $b(t)$ is now much more efficiently stored in the form chosen for the columns of $\boldsymbol{A}$ and the $n = 11$ scaling coefficients that are the components of $\boldsymbol{x}$.

```
octave] m=1000; h=2*pi/m; j=1:m;
octave] t(j)=(j-1)*h; t=transpose(t);
octave] n=5; A=[];
octave] for k=1:n
          A = [A sin(k*t)];
        end
octave] bt=t.*(pi-t).*(2*pi-t);
octave] x=A\bt;
octave] b=A*x;
octave] s=50; i=1:s:m;
        ts=t(i); bs=bt(i);
        plot(ts,bs,'ok',t,b,'r');
octave] print -depsc L01Fig02.eps
octave] close;
```

```
octave]
```

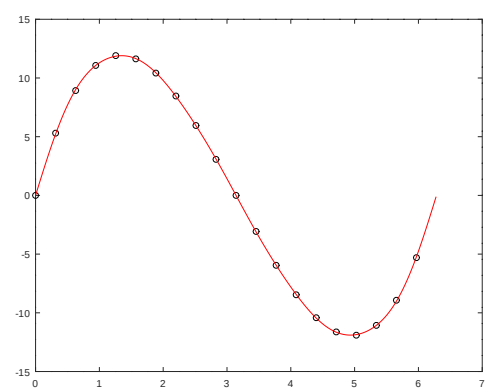

Figure 1.3. Comparison of least squares approximation (red line) with samples of exact function $b(t)$.

5. Vectors and matrice in data science

The above examples highlight some essential aspects of linear algebra in the context of data science applications.

- Vectors organize information that cannot be expressed as a single number and for which there exists a concept of scaling and addition.

- Matrices group together multiple vectors.

- The matrix-vector product expresses a linear combination of the column vectors of the matrix.

- Solving a linear system $\boldsymbol{Ax} = \boldsymbol{b} = \boldsymbol{Ib}$, to find $\boldsymbol{x} \in \mathbb{R}^m$ for given $\boldsymbol{b} \in \mathbb{R}^m$, re-expresses the linear combination

  $$\boldsymbol{b} = b_1\boldsymbol{e}_1 + \cdots + b_m\boldsymbol{e}_m, \boldsymbol{I} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \dots & \boldsymbol{e}_m \end{bmatrix},$$

  as another linear combination

  $$\boldsymbol{b} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \dots x_n\boldsymbol{a}_n, \boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}.$$

  For certain problems the linear combination $\boldsymbol{Ax}$ might be more insightful, but the above transformation is information-preserving, with $\boldsymbol{b}, \boldsymbol{x}$ both having the same number of components.

- Finding the best approximation of some given $\boldsymbol{b} \in \mathbb{R}^m$ by a linear combination $\boldsymbol{Ax}$ of the n column vectors of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is known as a least squares problem and transforms the information from the m components of $\boldsymbol{b}$ into n components of $\boldsymbol{x}$, and knowledge of the form of the column vectors. If $m > n$ and the form of the columns of $\boldsymbol{A}$ can be succintly stated, the transformation compresses information.

Data science seeks to extract regularity directly from available data, not necessarily invoking any additional hypotheses. The typical scenario is that immense amounts of data are available, with limited capability of human analysis. In this context it is apparent that the least squares problem is of greater interest than solving a linear system with a square matrix. It should also be clear that while computation by hand of small examples is useful to solidify theroretical concepts, it is essential to become proficient in the use of software that can deal with large data sets, such as Octave.

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 $X \times Y = \{(x, y) \mid x \in X, y \in Y\}$, 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 $X \times Y$, $R \subseteq X \times Y$.*

Similar to the difficulties encountered in attempting rigorous definition of a natural number, careful parsing of the above definition also would reveal self-references since the member of symbol $\in$, and the subset of symbol $\subseteq$ are both themselves examples of relations. As before, this is set aside to concetrate on the key concept of *associating an input $x \in X$ with an output $y \in Y$*. Associating an output to an input is also useful, leading to the definition of an *inverse relation* as $R^{-1} \subseteq Y \times X$, $R^{-1} = \{(y, x) \mid (x, y) \in R\}$. 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.

Computers can be programmed to work not only with numbers as Octave does, but also with general symbols as exemplified by another freely available software package called Maxima. Most data science applications involve numerical computation, but some knowledge of symbolic computation is also useful, as when working with sets that often arises in data classification. The colon symbol denotes assignment in Maxima, and sets can be defined using curly braces with automatic elimination of repeated elements. All common set manipulations are provided, such as the Cartesian product $\times$, and element of $\in$. operations.

```
%i1] X: {a,b,c,b,a}
```

(%o1) $\{a, b, c\}$

```
%i2] Y: {alpha,beta,gamma}
```

(%o2) $\{\alpha, \beta, \gamma\}$

```
%i3] XxY: cartesian_product(X,Y)
```

(%o3) $\{[a, \alpha], [a, \beta], [a, \gamma], [b, \alpha], [b, \beta], [b, \gamma], [c, \alpha], [c, \beta], [c, \gamma]\}$

```
%i4] [elementp([a,alpha],XxY), elementp([alpha,a],XxY)]
```

(%o4) $[\mathbf{true}, \mathbf{false}]$

```
%i5]
```

Associate the first three Latin and Greek letters by defining $R = \{(a, \alpha), (b, \beta), (c, \gamma)\}$. This is a relation between $X = \{a, b, c\}$ and $Y = \{\alpha, \beta, \gamma\}$ since it is a subset of $X \times Y$, which can be checked by defining a function that checks whether some $r \in R$ is also an element of $X \times Y$. Maxima functions are defined using the `:=` operator, and `map` applies a function to all elements of a set.

```
%i10] R: {[a,alpha],[b,beta],[c,gamma]}$
      verifyXxY(r):= elementp(r,XxY)$
      map(verifyXxY,R)
```

(%o12) $\{\mathbf{true}\}$

```
%i13]
```

Suppose that the Greek alphabet ordering is not known, and a might conceivably be associated to any of α, β, γ. This defines another relation $S = \{(a, \alpha), (a, \beta), (a, \gamma)\}$. Finally consider possible reorderings of the Greek alphabet, formulated as relationships between Y and itself, with $P = \{(\alpha, \beta), (\beta, \gamma), (\gamma, \alpha)\}$ and $I = \{(\alpha, \alpha), (\beta, \beta), (\gamma, \gamma)\}$ two such possible reorderings. The relations R, S, P, I defined here will be used to exemplify various properties below.

```
%i13] S: {[a,alpha],[a,beta],[a,gamma]}$
      YxY: cartesian_product(Y,Y)$ verifyYxY(r):= elementp(r,YxY)$
      P: {[alpha,beta],  [beta,gamma], [gamma,alpha]}$
      I: {[alpha,alpha], [beta,beta],  [gamma,gamma]}$
      [map(verifyXxY,S), map(verifyYxY,P), map(verifyYxY,I)]
```

(%o18) $[\{\mathbf{true}\},\{\mathbf{true}\},\{\mathbf{true}\}]$

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. Among the above-defined relations $P, I \subseteq Y \times Y$ are homogeneous, while $R, S \subseteq X \times Y$ are not. 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 $G \subseteq \mathbb{R} \times \mathbb{R}$, $G = L^{-1}$, $a < b \Longrightarrow b > a$. Homogeneous relations $H \subseteq A \times A$ are classified according to the following criteria.

Reflection. Relation H is reflexive if $(a, a) \in H$ for any $a \in A$. The equality relation $E \subseteq \mathbb{R} \times \mathbb{R}$ is reflexive, $\forall a \in A$, $a = a$, the less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$ is not, $1 \in R, 1 \not< 1$.

Symmetry. Relation H is symmetric if $(a, b) \in H$ implies that $(b, a) \in H$, $(a, b) \in H \Longrightarrow (b, a) \in H$. The equality relation $E \subseteq \mathbb{R} \times \mathbb{R}$ is symmetric, $a = b \Longrightarrow b = a$, the less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$ is not, $a < b \not\Longrightarrow b < a$.

Anti-symmetry. Relation H is anti-symmetric if $(a, b) \in H$ for $a \neq b$, then $(b, a) \notin H$. The less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$ is antisymmetric, $a < b \Longrightarrow b \not< a$.

Transitivity. Relation H is transitive if $(a, b) \in H$ and $(b, c) \in H$ implies $(a, c) \in H$. for any $a \in A$. The equality relation $E \subseteq \mathbb{R} \times \mathbb{R}$ is transitive, $a = b \wedge b = c \Longrightarrow a = c$, as is the less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$, $a < b \wedge b < c \Longrightarrow a < c$.

Certain combinations of properties often arise. A homogeneous relation that is reflexive, symmetric, and transitive is said to be an *equivalence relation*. Equivalence relations include equality among the reals, or congruence among triangles. A homogeneous relation that is reflexive, anti-symmetric and transitive is a *partial order relation*, such as the less than or equal relation between reals. Finally, a homogeneous relation that is anti-symmetric and transitive is an *order relation*, such as the less than relation between reals.

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 $f : X \longrightarrow Y$, with $F = \{(x, f(x)) \mid x \in X, f(x) \in Y\}$ and the property that $(x, y) \in F \Longrightarrow y = f(x)$. 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 $y = f(x)$. The previously defined relations R, P, I are functions but $S = \{(a, \alpha), (a, \beta), (a, \gamma)\}$ is not. 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 $\cos : \mathbb{R} \longrightarrow [-1, 1]$, $\sin : \mathbb{R} \longrightarrow [-1, 1]$ 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 $\exp : \mathbb{R} \longrightarrow \mathbb{R}$, $\log : (0, \infty) \longrightarrow \mathbb{R}$, as well as power and logarithm functions in some other base a;

- polynomial functions $p_n: \mathbb{R} \longrightarrow \mathbb{R}$, 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 Octave, and when given a vector argument return the function applied to each vector component.

```
octave] disp(cos(0:pi/4:pi))
```

```
   1.0000e+00   7.0711e-01   6.1232e-17  -7.0711e-01  -1.0000e+00
```

```
octave] y=log2(1:8); disp(y)
```

```
   0.00000   1.00000   1.58496   2.00000   2.32193   2.58496   2.80735   3.00000
```

```
octave] disp(pow2(y))
```

```
   1.0000   2.0000   3.0000   4.0000   5.0000   6.0000   7.0000   8.0000
```

```
octave] a=[1 0 -1]; x=-2:2; y=polyval(a,x); disp(y)
```

```
   3   0  -1   0   3
```

```
octave]
```

As seen previously, a Euclidean space $E_m = (\mathbb{R}^m, \mathbb{R}, +, \cdot)$ can be used to suggest properties of more complex spaces such as the vector space of continuous functions $\mathscr{C}^0(\mathbb{R})$. 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 $v: \{1, 2, \ldots, m\} \longrightarrow \mathbb{R}$, 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 $f: V \longrightarrow S$, or vectors from some other vector space $\mathscr{W} = (W, S, +, \cdot)$, $\boldsymbol{g}: V \longrightarrow W$. The codomain of a vector-valued function might be the same set of vectors as its domain, $\boldsymbol{h}: V \longrightarrow V$. 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 $f: V \longrightarrow S$ 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*

$$f(a\boldsymbol{u} + b\boldsymbol{v}) = af(\boldsymbol{u}) + bf(\boldsymbol{v}). \tag{1.5}$$

Many different functionals may be defined on a vector space $\mathscr{V} = (V, S, +, \cdot)$, and an insightful alternative description is provided by considering the set of all linear functionals, that will be denoted as $V^* = \{f \mid f: V \longrightarrow S\}$. These can be organized into another vector space $\mathscr{V}^* = (V^*, S, +, \cdot)$ with vector addition of linear functionals $f, g \in V^*$ and scaling by $a \in S$ defined by

$$(f + g)(\boldsymbol{u}) = f(\boldsymbol{u}) + g(\boldsymbol{u}), (af)(\boldsymbol{u}) = af(\boldsymbol{u}), \boldsymbol{u} \in V. \tag{1.6}$$

DEFINITION. (DUAL VECTOR SPACE) . *For some vector space $\mathscr{V}$, the vector space of linear functionals $\mathscr{V}^*$ is called the dual vector space.*

As is often the case, the above abstract definition can better be understood by reference to the familiar case of Euclidean space. Consider $\mathscr{R}_2 = (\mathbb{R}^2, \mathbb{R}, +, \cdot)$, the set of vectors in the plane with $\boldsymbol{x} \in \mathbb{R}^2$ the position vector from the origin $(0,0)$ to point X in the plane with coordinates (x_1, x_2). One functional from the dual space $\mathscr{R}_2^*$ is $f_2(\boldsymbol{x}) = x_2$, i.e., taking the second coordinate of the position vector. The linearity property is readily verified. For $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{R}_2$, $a, b \in \mathbb{R}$,

$$f_2(a\boldsymbol{x} + b\boldsymbol{y}) = ax_2 + by_2 = af_2(\boldsymbol{x}) + bf_2(\boldsymbol{y}).$$

Given some constant value $h \in \mathbb{R}$, the curves within the plane defined by $f_2(\boldsymbol{x}) = h$ are called the *contour lines* or *level sets* of f_2. Several contour lines and position vectors are shown in Figure ?. The utility of functionals and dual spaces can be shown by considering a simple example from physics. Assume that x_2 is the height above ground level and a vector $\boldsymbol{x}$ is the displacement of a body of mass m in a gravitational field. The mechanical work done to lift the body from ground level to height h is $W = mgh$ with g the gravitational acceleration. The mechanical work is the same for all displacements $\boldsymbol{x}$ that satisfy the equation $f_2(\boldsymbol{x}) = h$. The work expressed in units $mg\Delta h$ can be interpreted as the number of contour lines $f_2(\boldsymbol{x}) = n\Delta h$ intersected by the displacement vector $\boldsymbol{x}$. This concept of duality between vectors and scalar-valued functionals arises throughout mathematics, the physical and social sciences and in data science. The term "duality" itself comes from geometry. A point X in $\mathbb{R}^2$ with coordinates (x_1, x_2) can be defined either as the end-point of the position vector $\boldsymbol{x}$, or as the intersection of the contour lines of two funtionals $f_1(\boldsymbol{x}) = x_1$ and $f_2(\boldsymbol{x}) = x_2$. Either geometric description works equally well in specifying the position of X, so it might seem redundant to have two such procedures. It turns out though that many quantities of interest in applications can be defined through use of both descriptions, as shown in the computation of mechanical work in a gravitational field.

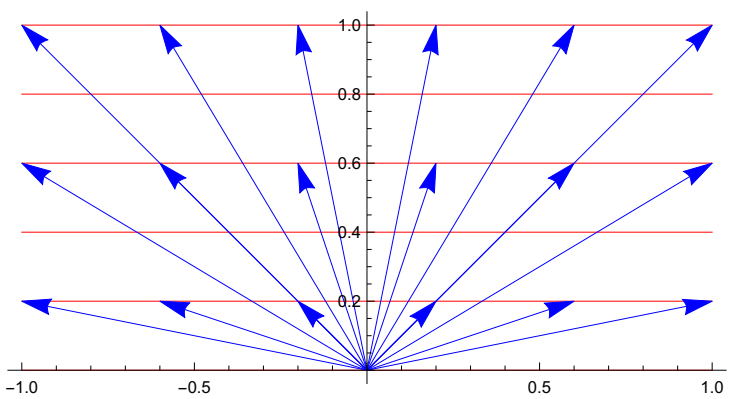

Figure 1.4. Vectors in E_2 and contour lines of the functional $f(\boldsymbol{x}) = x_2$

1.4. Linear mappings

Consider now functions $\boldsymbol{f}: V \longrightarrow W$ 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 $\boldsymbol{f}: V \longrightarrow W$, from vector space $\mathscr{V} = (V, S, +, \cdot)$ to vector space $\mathscr{W} = (W, S, \oplus, \odot)$ is called a linear mapping if for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ and any two scalars $a, b \in S$*

$$\boldsymbol{f}(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{f}(\boldsymbol{u}) + b\boldsymbol{f}(\boldsymbol{v}). \tag{1.7}$$

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, and called *homomorphisms* in mathematics. The codomain of a linear mapping might be the same as the domain in which case the mapping is said to be an *endomorphism*.

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

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} = x_1\,\boldsymbol{a}_1 + \cdots + x_n\,\boldsymbol{a}_n,$$

with $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ the columns of the matrix, $\boldsymbol{A} = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \ldots\ \ \boldsymbol{a}_n\ ]$. This is a linear mapping between the real spaces $\mathscr{R}_m$, $\mathscr{R}_n$, $\boldsymbol{f}: \mathbb{R}^m \longrightarrow \mathbb{R}^n$, and indeed any linear mapping between real spaces can be given as a matrix-vector product.

2. Measurements

Vectors within the real space $\mathscr{R}_m$ can be completely specified by m real numbers, even though m is large in many realistic applications. A vector within $\mathscr{C}^0(\mathbb{R})$, i.e., a continuous function defined on the reals, cannot be so specified since it would require an infinite, non-countable listing of function values. In either case, 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.

2.1. Equivalence classes

Many classification criteria are scalars, defined as a scalar-valued function $f: \mathscr{V} \longrightarrow S$ on a vector space, $\mathscr{V} = (V, S, +, \cdot)$. The most common criteria are inspired by experience with Euclidean space. In a Euclidean-Cartesian model $(\mathbb{R}^2, \mathbb{R}, +, \cdot)$ of the geometry of a plane Π, a point $O \in \Pi$ is arbitrarily chosen to correspond to the zero vector $\boldsymbol{0} = [\ 0\ \ 0\ ]^T$, along with two preferred vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$ grouped together into the identity matrix $\boldsymbol{I}$. The position of a point $X \in \Pi$ with respect to O is given by the linear combination

$$\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} + \boldsymbol{0} = [\ \boldsymbol{e}_1\ \ \boldsymbol{e}_2\ ] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\,\boldsymbol{e}_1 + x_2\,\boldsymbol{e}_2\,.$$

Several possible classifications of points in the plane are depicted in Figure ?: lines, squares, circles. Intuitively, each choice separates the plane into subsets, and a given point in the plane belongs to just one in the chosen family of subsets. A more precise characterization is given by the concept of a partition of a set.

DEFINITION. (PARTITION) . *A partition of a set is a grouping of its elements into non-empty subsets such that every element is included in exactly one subset.*

In precise mathematical terms, a partition of set S is $P = \{S_i \mid S_i \subset P, S_i \neq \emptyset, i \in I\}$ such that $\forall x \in S, \exists!\, j \in I$ for which $x \in S_j$. Since there is only one set ($\exists!$ signifies "exists and is unique") to which some given $x \in S$ belongs, the subsets S_i of the partition P are disjoint, $i \neq j \Longrightarrow S_i \cap S_j = \emptyset$. The subsets S_i are labeled by i within some index set I. The index set might be a subset of the naturals, $I \subset \mathbb{N}$ in which case the partition is countable, possibly finite. The partitions of the plane suggested by Figure ? are however indexed by a real-valued label, $i \in \mathbb{R}$ with $I \subset \mathbb{R}$.

A technique which is often used to generate a partition of a vector space $\mathscr{V} = (V, S, +, \cdot)$ is to define an equivalence relation between vectors, $H \subseteq V \times V$. For some element $\boldsymbol{u} \in V$, the *equivalence class* of $\boldsymbol{u}$ is defined as all vectors $\boldsymbol{v}$ that are equivalent to $\boldsymbol{u}$, $\{\boldsymbol{v} \mid (\boldsymbol{u}, \boldsymbol{v}) \in H\}$. The set of equivalence classes of is called the *quotient set* and denoted as V/H, and the quotient set is a partition of V. Figure ? depicts four different partitions of the plane. These can be interpreted geometrically, such as parallel lines or distance from the origin. With wider implications for linear algebra, the partitions can also be given in terms of classification criteria specified by functions.

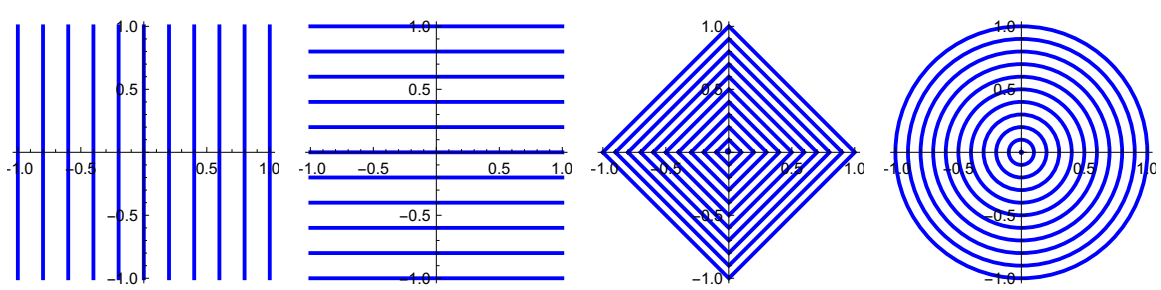

Figure 1.5. Equivalence classes within the plane

2.2. Norms

The partition of $\mathbb{R}^2$ by circles from Figure 1.5 is familiar; the equivalence classes are sets of points whose position vector has the same size, $\{\boldsymbol{x} = [\ x_1\ \ x_2\ ]^T \mid (x_1^2 + x_2^2)^{1/2} = r\}$, or is at the same distance from the origin. Note that familiarity with Euclidean geometry should not obscure the fact that some other concept of distance might be induced by the data. A simple example is statement of walking distance in terms of city blocks, in which the distance from a starting point to an address $x_1 = 3$ blocks east and $x_2 = 4$ blocks north is $x_1 + x_2 = 7$ city blocks, not the Euclidean distance $(x_1^2 + x_2^2)^{1/2} = 5$ since one cannot walk through the buildings occupying a city block.

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 $\boldsymbol{v} \in V$ 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 $\mathbb{R}$, $f: V \longrightarrow \mathbb{R}$. On the real line the point at coordinate x is at distance $|x|$ from the origin, and to mimic this usage the norm of $\boldsymbol{v} \in V$ is denoted as $\|\boldsymbol{v}\|$, leading to the definition of a function $\|\ \|: V \longrightarrow \mathbb{R}_+$, $\mathbb{R}_+ = \{a \mid a \in \mathbb{R}, a \geqslant 0\}$.

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, $\|\boldsymbol{v}\| = 0 \Longrightarrow \boldsymbol{v} = \mathbf{0}$. 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 another branch of mathematics known as 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*, $\|\boldsymbol{u} + \boldsymbol{v}\| \leqslant \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ for $\boldsymbol{u}, \boldsymbol{v} \in V$.

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

1. $\|\boldsymbol{v}\| = 0 \Longrightarrow \boldsymbol{v} = \mathbf{0}$;
2. $\|a\boldsymbol{u}\| = |a|\ \|\boldsymbol{u}\|$;
3. $\|\boldsymbol{u} + \boldsymbol{v}\| \leqslant \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.

Note that the norm is a functional, but the triangle inequality implies that it is not generally a linear functional. Returning to Figure 1.5, consider the functions $f_i: \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ defined for $\boldsymbol{x} = [\ x_1\ \ x_2\ ]^T$ through values

$$f_1(\boldsymbol{x}) = |x_1|,\ f_2(\boldsymbol{x}) = |x_2|,\ f_3(\boldsymbol{x}) = |x_1| + |x_2|,\ f_4(\boldsymbol{x}) = (|x_1|^2 + |x_2|^2)^{1/2}.$$

Sets of constant value of the above functions are also equivalence classes induced by the equivalence relations E_i for $i = 1, 2, 3, 4$.

1. $f_1(\boldsymbol{x}) = c \Longrightarrow |x_1| = c$, $E_1 = \{(\boldsymbol{x}, \boldsymbol{y}) \mid f_1(\boldsymbol{x}) = f_1(\boldsymbol{y}) \Longleftrightarrow |x_1| = |y_1|\} \subseteq \mathbb{R}^2 \times \mathbb{R}^2$;
2. $f_2(\boldsymbol{x}) = c \Longrightarrow |x_2| = c$, $E_2 = \{(\boldsymbol{x}, \boldsymbol{y}) \mid f_2(\boldsymbol{x}) = f_2(\boldsymbol{y}) \Longleftrightarrow |x_2| = |y_2|\} \subseteq \mathbb{R}^2 \times \mathbb{R}^2$;
3. $f_3(\boldsymbol{x}) = c \Longrightarrow |x_1| + |x_2| = c$, $E_3 = \{(\boldsymbol{x}, \boldsymbol{y}) \mid f_3(\boldsymbol{x}) = f_3(\boldsymbol{y}) \Longleftrightarrow |x_1| + |x_2| = |y_1| + |y_2|\} \subseteq \mathbb{R}^2 \times \mathbb{R}^2$;
4. $f_4(\boldsymbol{x}) = c \Longrightarrow (|x_1|^2 + |x_2|^2)^{1/2} = c$, $E_4 = \{(\boldsymbol{x}, \boldsymbol{y}) \mid f_4(\boldsymbol{x}) = f_4(\boldsymbol{y}) \Longleftrightarrow (|x_1|^2 + |x_2|^2)^{1/2} = (|y_1|^2 + |y_2|^2)^{1/2}\} \subseteq \mathbb{R}^2 \times \mathbb{R}^2$.

These equivalence classes correspond to the vertical lines, horizontal lines, squares, and circles of Figure 1.5. Not all of the functions f_i are norms since $f_1(\boldsymbol{x})$ is zero for the non-null vector $\boldsymbol{x} = [\ 0\ \ 1\ ]^T$, and $f_2(\boldsymbol{x})$is zero for the non-null vector $\boldsymbol{x} = [\ 1\ \ 0\ ]^T$. The functions f_3 and f_4 are indeed norms, and specific cases of the following general norm.

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* $p \geqslant 1$ *is the function* $\| \ \|_p$: $V \longrightarrow \mathbb{R}_+$ *with values* $\|\boldsymbol{x}\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_m|^p)^{1/p}$, *or*

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p} \text{ for } \boldsymbol{x} \in \mathbb{R}^m. \tag{1.8}$$

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

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

This also works for vectors with equal components, since the fact that the number of components is finite while $p \longrightarrow \infty$ can be used as exemplified for $\boldsymbol{x} = [\ a \ \ a \ \ \ldots \ \ a\ ]^T$, by $\|\boldsymbol{x}\|_p = (m\,|a|^p)^{1/p} = m^{1/p}\,|a|$, with $m^{1/p} \longrightarrow 1$.

Note that the Euclidean norm corresponds to $p = 2$, and is often called the 2-norm. The analogy between vectors and functions can be exploited to also define a p-norm for $\mathscr{C}^0[a,b] = (C([a,b]), \mathbb{R}, +, \cdot)$, the vector space of continuous functions defined on $[a,b]$.

DEFINITION. (p-NORM IN $\mathscr{C}^0[a,b]$) . *The p-norm on the vector space of continuous functions* $\mathscr{C}^0[a,b]$ *for* $p \geqslant 1$ *is the function* $\| \ \|_p$: $V \longrightarrow \mathbb{R}_+$ *with values*

$$\|f\|_p = \left(\int_a^b |f(x)|^p \,\mathrm{d}x\right)^{1/p}, \text{for } f \in C[a,b]. \tag{1.9}$$

The integration operation $\int_a^b$ can be intuitively interpreted as the value of the sum $\sum_{i=1}^m$ from equation (1.8) for very large m and very closely spaced evaluation points of the function $f(x_i)$, for instance $|x_{i+1} - x_i| = (b-a)/m$. An inf-norm can also be define for continuous functions by

$$\|f\|_\infty = \sup_{x \in [a,b]} |f(x)|,$$

where sup, the supremum operation can be intuitively understood as the generalization of the max operation over the countable set $\{1, 2, \ldots, m\}$ to the uncountable set $[a,b]$.

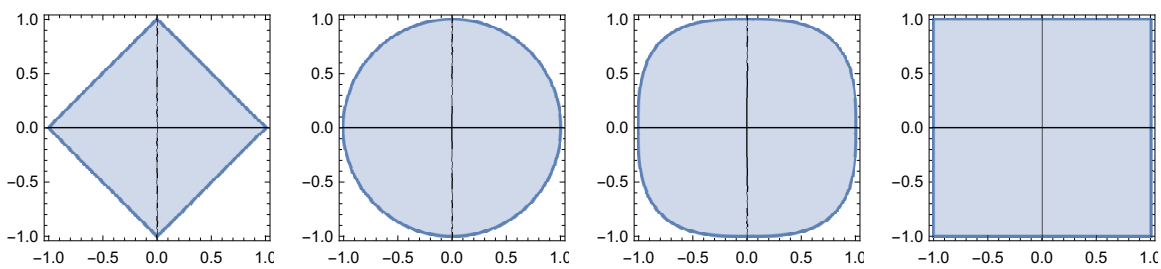

Figure 1.6. Regions within $\mathbb{R}^2$ for which $\|\boldsymbol{x}\|_p \leqslant 1$, for $p = 1, 2, 3, \infty$.

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 Octave 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.

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

```
   1.7321   1.7321
```

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

```
   3   3
```

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

```
   4   4
```

```
octave] disp([norm(x,1) norm(x,2) norm(x,4) norm(x,8) norm(x,16) norm(x,inf)])
    4.0000   2.0000   1.4142   1.1892   1.0905   1.0000
octave]
```

2.3. 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, but before a formal definition some intuitive understanding is sought by considering vectors and functionals in the plane, as depicted in Figure ?. Consider a position vector $\boldsymbol{x} = [\ x_1\ \ x_2\ ]^T \in \mathbb{R}^2$ and the previously-encountered linear functionals

$$f_1, f_2 \colon \mathbb{R}^2 \longrightarrow \mathbb{R},\ f_1(\boldsymbol{x}) = x_1,\ f_2(\boldsymbol{x}) = x_2.$$

The x_1 component of the vector $\boldsymbol{x}$ can be thought of as the number of level sets of f_1 times it crosses; similarly for the x_2 component. A convenient labeling of level sets is by their normal vectors. The level sets of f_1 have normal $\boldsymbol{e}_1^T = [\ 1\ \ 0\ ]$, and those of f_2 have normal vector $\boldsymbol{e}_2^T = [\ 0\ \ 1\ ]$. Both of these can be thought of as matrices with two columns, each containing a single component. The products of these matrices with the vector $\boldsymbol{x}$ gives the value of the functionals f_1, f_2

$$\boldsymbol{e}_1^T \boldsymbol{x} = [\ 1\ \ 0\ ] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \cdot x_1 + 0 \cdot x_2 = x_1 = f_1(\boldsymbol{x}),$$

$$\boldsymbol{e}_2^T \boldsymbol{x} = [\ 0\ \ 1\ ] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \cdot x_1 + 1 \cdot x_2 = x_1 = f_2(\boldsymbol{x}).$$

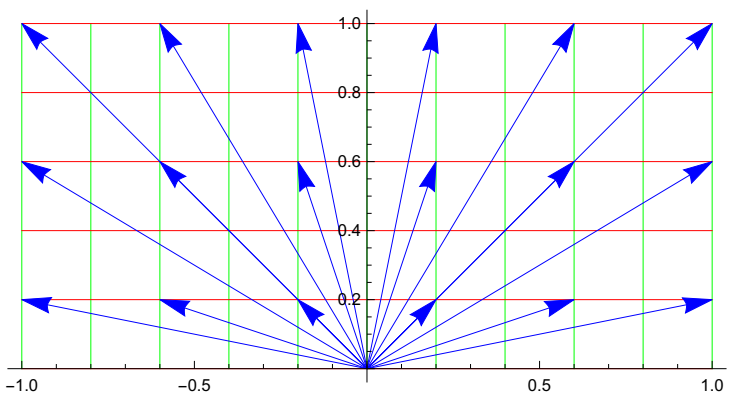

Figure 1.7. Euclidean space E_2 and its dual E_2^*.

In general, any linear functional f defined on the real space $\mathscr{R}_m$ can be labeled by a vector

$$\boldsymbol{a}^T = [\ a_1\ \ a_2\ \ \ldots\ \ a_m\ ],$$

and evaluated through the matrix-vector product $f(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}$. This suggests the definition of another function s: $\mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}$,

$$s(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}.$$

The function s is called an *inner product*, has two vector arguments from which a matrix-vector product is formed and returns a scalar value, hence is also called a *scalar product*. The definition from an Euclidean space can be extended to general vector spaces. For now, consider the field of scalars to be the reals $S = \mathbb{R}$.

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

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

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

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

The inner product $s(\boldsymbol{a}, \boldsymbol{x})$ returns the number of level sets of the functional labeled by $\boldsymbol{a}$ crossed by the vector $\boldsymbol{x}$, and this interpretation underlies many applications in the sciences as in the gravitational field example above. Inner products also provide a procedure to evaluate geometrical quantities and relationships.

Vector norm. In $\mathscr{R}_m$ the number of level sets of the functional labeled by $\boldsymbol{x}$ crossed by $\boldsymbol{x}$ itself is identical to the square of the 2-norm

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

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

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

A real space together with the scalar product $s(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y}$ and induced norm $\| \boldsymbol{x} \| = s(\boldsymbol{x}, \boldsymbol{x})^{1/2}$ 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 $\boldsymbol{x} = [\ x_1\ \ x_2\ ]^T$. The inner product

$$\boldsymbol{e}_1^T\boldsymbol{x} = [\ 1\ \ 0\ ]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \cdot x_1 + 0 \cdot x_2 = r\cos\theta,$$

is seen to contain information on the relative orientation of $\boldsymbol{x}$ with respect to $\boldsymbol{e}_1$. In general, the angle θ between two vectors $\boldsymbol{x}, \boldsymbol{y}$ with any vector space with a scalar product can be defined by

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

which becomes

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

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

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 $\boldsymbol{x}, \boldsymbol{y} \in V$ are orthogonal if $s(\boldsymbol{x}, \boldsymbol{y}) = 0$. In $\mathscr{E}_m$ vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ are orthogonal if $\boldsymbol{x}^T\boldsymbol{y} = 0$.

3. Linear mapping composition

3.1. Matrix-matrix product

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

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

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

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

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

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{g}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{A}\boldsymbol{x}.$$

Note that the intemediate vector $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$ is subsequently multiplied by the matrix $\boldsymbol{B}$. The composite function $\boldsymbol{h}$ is itself a linear mapping

$$\boldsymbol{h}(a\boldsymbol{x} + b\boldsymbol{y}) = \boldsymbol{B}\boldsymbol{A}(a\boldsymbol{x} + b\boldsymbol{y}) = \boldsymbol{B}\,(a\boldsymbol{A}\boldsymbol{x} + b\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{B}(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{B}\boldsymbol{u} + b\boldsymbol{B}\boldsymbol{v} = a\boldsymbol{B}\boldsymbol{A}\boldsymbol{x} + b\boldsymbol{B}\boldsymbol{A}\boldsymbol{y} = a\boldsymbol{h}(\boldsymbol{x}) + b\boldsymbol{h}(\boldsymbol{y}),$$

so it also can be expressed a matrix-vector multiplication

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{C}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{A}\boldsymbol{x}. \tag{1.10}$$

Using the above, $\boldsymbol{C}$ is defined as the product of matrix $\boldsymbol{B}$ with matrix $\boldsymbol{A}$

$$\boldsymbol{C} = \boldsymbol{B}\boldsymbol{A}.$$

The columns of $\boldsymbol{C}$ can be determined from those of $\boldsymbol{A}$ by considering the action of $\boldsymbol{h}$ on the the column vectors of the identity matrix $\boldsymbol{I} = [\ \boldsymbol{e}_1\ \ \boldsymbol{e}_2\ \ \dots\ \ \boldsymbol{e}_n\ ] \in \mathbb{R}^{n\times n}$. First, note that

$$\boldsymbol{A}\boldsymbol{e}_j = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \dots\ \ \boldsymbol{a}_n\ ]\begin{bmatrix}1\\0\\\vdots\\\vdots\\0\end{bmatrix} = \boldsymbol{a}_1, \dots, \boldsymbol{A}\boldsymbol{e}_j = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \dots\ \ \boldsymbol{a}_n\ ]\begin{bmatrix}0\\\vdots\\1\\\vdots\\0\end{bmatrix} = \boldsymbol{a}_j, \boldsymbol{A}\boldsymbol{e}_n = [\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \dots\ \ \boldsymbol{a}_n\ ]\begin{bmatrix}0\\\vdots\\\vdots\\0\\1\end{bmatrix} = \boldsymbol{a}_n. \tag{1.11}$$

The above can be repeated for the matrix $\boldsymbol{C} = [\ \boldsymbol{c}_1\ \ \boldsymbol{c}_2\ \ \dots\ \ \boldsymbol{c}_n\ ]$ giving

$$\boldsymbol{h}(\boldsymbol{e}_1) = \boldsymbol{C}\boldsymbol{e}_1 = \boldsymbol{c}_1, \dots, \boldsymbol{h}(\boldsymbol{e}_j) = \boldsymbol{C}\boldsymbol{e}_j = \boldsymbol{c}_j, \dots, \boldsymbol{h}(\boldsymbol{e}_n) = \boldsymbol{C}\boldsymbol{e}_n = \boldsymbol{c}_n. \tag{1.12}$$

Combining the above equations leads to $\boldsymbol{c}_j = \boldsymbol{B}\boldsymbol{a}_j$, or

$$\boldsymbol{C} = [\ \boldsymbol{c}_1\ \ \boldsymbol{c}_2\ \ \dots\ \ \boldsymbol{c}_n\ ] = \boldsymbol{B}\,[\ \boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \dots\ \ \boldsymbol{a}_n\ ].$$

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

$$\boldsymbol{C} = [\ \boldsymbol{c}_1\ \ \boldsymbol{c}_2\ \ \dots\ \ \boldsymbol{c}_n\ ] = [\boldsymbol{B}\ \boldsymbol{a}_1\ \ \boldsymbol{B}\boldsymbol{a}_2\ \ \dots\ \ \boldsymbol{B}\boldsymbol{a}_n\ ]$$

Matrix-vector and matrix-matrix products are implemented in Octave, the above results can readily be verified.

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

```
A =

   1   3
   2   4
```

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

```
B =

  -1   2
   1  -2
   3   3
```

```
octave] C=B*A
```

```
C =

   3   5
  -3  -5
   9  21
```

```
octave] c1=B*a1; c2=B*a2; [c1 c2]
```

```
ans =

   3   5
  -3  -5
   9  21
```

```
octave]
```