

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 initial 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 $\mathcal{R}_2 = (\mathbb{R}^2, \mathbb{R}, +, \cdot)$ 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 concentrate 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 $R^{-1} \subseteq Y \times X$, $R^{-1} = \{(y, x) | (x, y) \in R\}$. 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) [true,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) {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) [{true},{true},{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 $G \subseteq \mathbb{R} \times \mathbb{R}$, $G = L^{-1}$, $a < b \Rightarrow b > a$. 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, $\forall a \in \mathbb{R}$, $a = a$, the less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$ is not, $1 \in \mathbb{R}$, $1 \not< 1$.

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

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, $a < b \Rightarrow b \not< a$.

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, $a = b \wedge b = c \Rightarrow a = c$, as is the less than relation $L \subseteq \mathbb{R} \times \mathbb{R}$, $a < b \wedge b < c \Rightarrow a < c$.

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 $f: X \rightarrow Y$, 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 $y = f(x)$. 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 $\cos: \mathbb{R} \rightarrow [-1, 1]$, $\sin: \mathbb{R} \rightarrow [-1, 1]$ 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 $\exp: \mathbb{R} \rightarrow \mathbb{R}$, $\log: (0, \infty) \rightarrow \mathbb{R}$, as well as power and logarithm functions in some other base a ;
- polynomial functions $p_n: \mathbb{R} \rightarrow \mathbb{R}$, defined by a succession of additions and multiplications

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

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 $\mathcal{C}^0(\mathbb{R})$. A construct that will be often used is to interpret a vector within E_m as a function, since $v \in \mathbb{R}^m$ with components $v = [v_1 \ v_2 \ \dots \ v_m]^T$ also defines a function $v: \{1, 2, \dots, m\} \rightarrow \mathbb{R}$, with values $v(i) = v_i$. As the number of components grows the function v can provide better approximations of some continuous function $f \in \mathcal{C}^0(\mathbb{R})$ through the function values $v_i = v(i) = f(x_i)$ at distinct sample points x_1, x_2, \dots, 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 $\mathcal{V} = (V, S, +, \cdot)$ that return either scalars $f: V \rightarrow S$, or vectors from some other vector space $\mathcal{W} = (W, S, +, \cdot)$, $g: V \rightarrow W$. The codomain of a vector-valued function might be the same set of vectors as its domain, $h: V \rightarrow V$. The fundamental operation within linear algebra is the linear combination $au + bv$ with $a, b \in S$, $u, v \in V$. A key aspect is to characterize how a function behaves when given a linear combination as its argument, for instance $f(au + bv)$ or $g(au + bv)$.

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 $\mathcal{V} = (V, S, +, \cdot)$ is a function from the set of vectors V to the set of scalars S of the vector space \mathcal{V} .

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

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v}). \quad (1)$$

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

$$(f + g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u}), (af)(\mathbf{u}) = af(\mathbf{u}), \mathbf{u} \in V. \quad (2)$$

DEFINITION. (DUAL VECTOR SPACE) . For some vector space \mathcal{V} , the vector space of linear functionals \mathcal{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 $\mathcal{R}_2 = (\mathbb{R}^2, \mathbb{R}, +, \cdot)$, the set of vectors in the plane with $\mathbf{x} \in \mathbb{R}^2$ 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 \mathcal{R}_2^* is $f_2(\mathbf{x}) = x_2$, i.e., taking the second coordinate of the position vector. The linearity property is readily verified. For $\mathbf{x}, \mathbf{y} \in \mathcal{R}_2$, $a, b \in \mathbb{R}$,

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

Given some constant value $h \in \mathbb{R}$, the curves within the plane defined by $f_2(\mathbf{x}) = h$ are called the *contour lines* or *level sets* of f_2 . Several contour lines and position vectors are shown in Figure 1. 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 \mathbf{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 \mathbf{x} that satisfy the equation $f_2(\mathbf{x}) = h$. The work expressed in units $mg\Delta h$ can be interpreted as the number of contour lines $f_2(\mathbf{x}) = n\Delta h$ intersected by the displacement vector \mathbf{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 \mathbf{x} , or as the intersection of the contour lines of two functionals $f_1(\mathbf{x}) = x_1$ and $f_2(\mathbf{x}) = x_2$. 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.

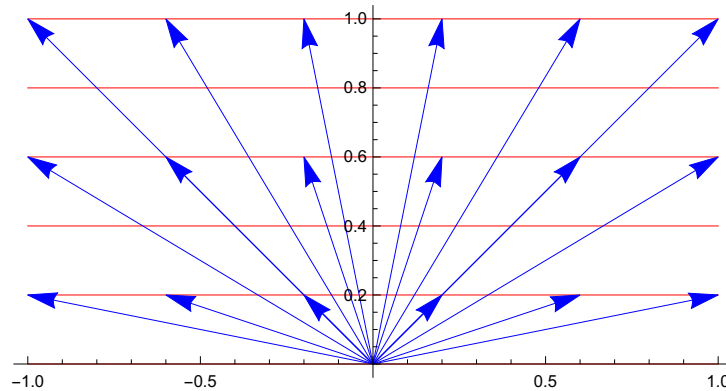


Figure 1. Vectors in E_2 and contour lines of the functional $f(\mathbf{x}) = x_2$

1.4. Linear mappings

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

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

$$f(a\mathbf{u} + b\mathbf{v}) = af(\mathbf{u}) + bf(\mathbf{v}). \quad (3)$$

The image of a linear combination $a\mathbf{u} + b\mathbf{v}$ through a linear mapping is another linear combination $af(\mathbf{u}) + bf(\mathbf{v})$, 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

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

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

2. Measurements

Vectors within the real space \mathcal{R}_m can be completely specified by m real numbers, even though m is large in many realistic applications. A vector within $\mathcal{C}^0(\mathbb{R})$, 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 $\mathcal{V} = (V, S, +, \cdot)$ 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 $f: \mathcal{V} \rightarrow S$ on a vector space, $\mathcal{V} = (V, S, +, \cdot)$. 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 $\mathbf{0} = [0 \ 0]^T$, along with two preferred vectors $\mathbf{e}_1, \mathbf{e}_2$ grouped together into the identity matrix \mathbf{I} . The position of a point $X \in \Pi$ with respect to O is given by the linear combination

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

Several possible classifications of points in the plane are depicted in Figure 2: 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, $i \neq j \Rightarrow S_i \cap S_j = \emptyset$. 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 2 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 $\mathcal{V} = (V, S, +, \cdot)$ is to define an equivalence relation between vectors, $H \subseteq V \times V$. For some element $\mathbf{u} \in V$, the *equivalence class* of \mathbf{u} is defined as all vectors \mathbf{v} that are equivalent to \mathbf{u} , $\{\mathbf{v} \mid (\mathbf{u}, \mathbf{v}) \in H\}$. 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 2 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.

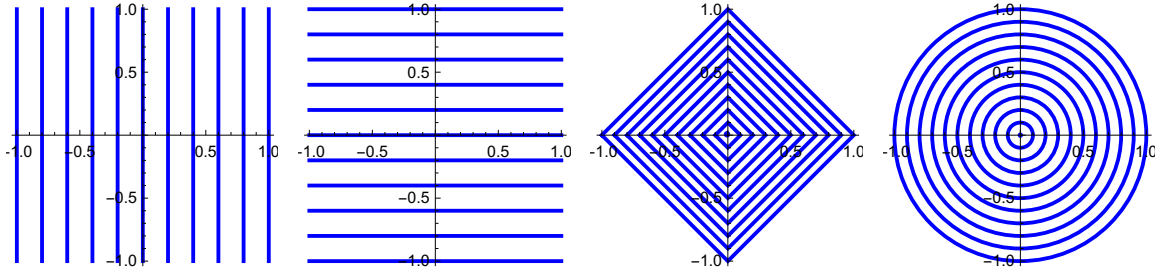


Figure 2. Equivalence classes within the plane

2.2. Norms

The partition of \mathbb{R}^2 by circles from Figure 2 is familiar; the equivalence classes are sets of points whose position vector has the same size, $\{\mathbf{x} = [x_1 \ x_2]^T \mid (x_1^2 + x_2^2)^{1/2} = r\}$, 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, $\mathcal{V} = (V, S, +, \cdot)$, 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 $\mathbf{v} \in V$ 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} , $f: V \rightarrow \mathbb{R}$. On the real line the point at coordinate x is at distance $|x|$ from the origin, and to mimic this usage the norm of $\mathbf{v} \in V$ is denoted as $\|\mathbf{v}\|$, leading to the definition of a function $\|\cdot\|: V \rightarrow \mathbb{R}_+$, $\mathbb{R}_+ = \{a \mid a \in \mathbb{R}, a \geq 0\}$.

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, $\|\mathbf{v}\| = 0 \Rightarrow \mathbf{v} = \mathbf{0}$. 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 algebra or the singular value decomposition encountered later.

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

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*, $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ for $\mathbf{u}, \mathbf{v} \in V$.

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

1. $\|\mathbf{v}\| = 0 \Rightarrow \mathbf{v} = \mathbf{0}$;
2. $\|a\mathbf{u}\| = |a|\|\mathbf{u}\|$;
3. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

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

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

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

These equivalence classes correspond to the vertical lines, horizontal lines, squares, and circles of Figure 2. Not all of the functions f_i are norms since $f_1(\mathbf{x})$ is zero for the non-null vector $\mathbf{x} = [0 \ 1]^T$, and $f_2(\mathbf{x})$ is zero for the non-null vector $\mathbf{x} = [1 \ 0]^T$. The functions f_3 and f_4 are indeed norms, and specific cases of the following general norm.

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

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^m |x_i|^p \right)^{1/p} \text{ for } \mathbf{x} \in \mathbb{R}^m. \tag{4}$$

Denote by x_i the largest component in absolute value of $\mathbf{x} \in \mathbb{R}^m$. 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

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

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

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 $\mathcal{C}^0[a, b] = (C([a, b]), \mathbb{R}, +, \cdot)$, the vector space of continuous functions defined on $[a, b]$.

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

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx \right)^{1/p}, \text{ for } f \in C[a, b]. \tag{5}$$

The integration operation \int_a^b can be intuitively interpreted as the value of the sum $\sum_{i=1}^m$ from equation (4) 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 $\{1, 2, \dots, m\}$ to the uncountable set $[a, b]$.

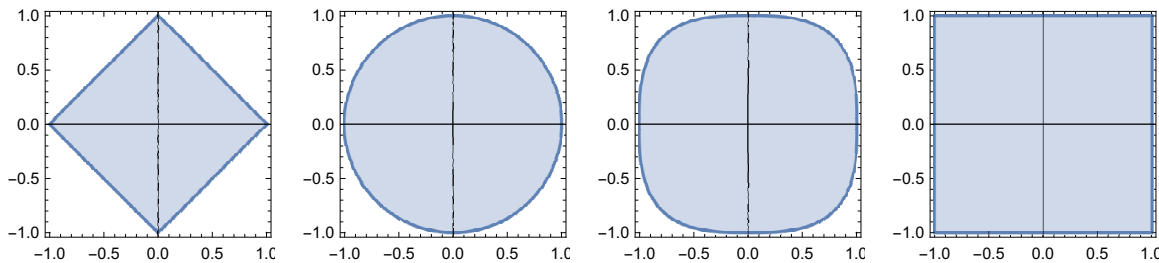


Figure 3. Regions within \mathbb{R}^2 for which $\|\mathbf{x}\|_p \leq 1$, for $p = 1, 2, 3, \infty$.

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 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 4. Consider a position vector $\mathbf{x} = [x_1 \ x_2]^T \in \mathbb{R}^2$ and the previously-encountered linear functionals

$$f_1, f_2: \mathbb{R}^2 \rightarrow \mathbb{R}, f_1(\mathbf{x}) = x_1, f_2(\mathbf{x}) = x_2.$$

The x_1 component of the vector \mathbf{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 $\mathbf{e}_1^T = [1 \ 0]$, and those of f_2 have normal vector $\mathbf{e}_2^T = [0 \ 1]$. 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 \mathbf{x} gives the value of the functionals f_1, f_2

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

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

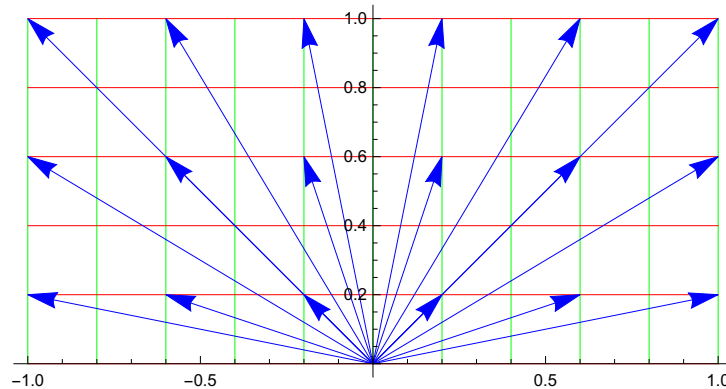


Figure 4. Euclidean space E_2 and its dual E_2^* .

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

$$\mathbf{a}^T = [a_1 \ a_2 \ \dots \ a_m],$$

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

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

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 $\mathcal{V} = (V, \mathbb{R}, +, \cdot)$ is a function $s: V \times V \rightarrow \mathbb{R}$ with properties

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

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

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

The inner product $s(\mathbf{a}, \mathbf{x})$ returns the number of level sets of the functional labeled by \mathbf{a} crossed by the vector \mathbf{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 \mathcal{F}_m the number of level sets of the functional labeled by \mathbf{x} crossed by \mathbf{x} itself is identical to the square of the 2-norm

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

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

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

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

Orientation. In \mathcal{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 $\mathbf{x} = [x_1 \ x_2]^T$. The inner product

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

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

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

which becomes

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

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

Orthogonality. In \mathcal{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 $\mathcal{V} = (V, \mathbb{R}, +, \cdot)$ with a scalar product by stating that $\mathbf{x}, \mathbf{y} \in V$ are orthogonal if $s(\mathbf{x}, \mathbf{y}) = 0$. In \mathcal{E}_m vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ are orthogonal if $\mathbf{x}^T \mathbf{y} = 0$.

3. Linear mapping composition

3.1. Matrix-matrix product

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

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

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

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

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

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

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

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

so it also can be expressed a matrix-vector multiplication

$$h(x) = Cx = BAx. \quad (6)$$

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

$$Ae_j = [a_1 \ a_2 \ \dots \ a_n] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = a_1, \dots, Ae_j = [a_1 \ a_2 \ \dots \ a_n] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_j, Ae_n = [a_1 \ a_2 \ \dots \ a_n] \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = a_n. \quad (7)$$

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

$$h(e_1) = Ce_1 = c_1, \dots, h(e_j) = Ce_j = c_j, \dots, h(e_n) = Ce_n = c_n. \quad (8)$$

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 ,

$$C = [c_1 \ c_2 \ \dots \ c_n] = [Ba_1 \ Ba_2 \ \dots \ Ba_n]$$

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]
```