A brief survey of linear algebra

Tom Carter

http://cogs.csustan.edu/~ tom/linear-algebra

Santa Fe Institute
Complex Systems Summer School

June, 2001

Our general topics: ←

(ex): exercises.

Why linear algebra ←

- Linear models of phenomena are pervasive throughout science. The techniques of linear algebra provide tools which are applicable in a wide variety of contexts.

  Beyond that, linear algebra courses are often the transition from lower division mathematics courses such as calculus, probability/statistics, and elementary differential equations, which typically focus on specific problem solving techniques, to the more theoretical axiomatic and proof oriented upper division mathematics courses.

  I am going to stay with a generally abstract, axiomatic presentation of the basics of linear algebra. (But I'll also try to provide some practical advice along the way . . . :-)

Vector spaces ←

- The first thing we need is a field $\mathbb{F}$ of coefficients for our vector space. The most frequently used fields are the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$. A **field** $\mathbb{F} = (\mathbb{F}, +, *, -, {}^{-1}, 0, 1)$ is a mathematical object consisting of a set of elements ($\mathbb{F}$), together with two binary operations ($+, *$), two unary operations ($-, {}^{-1}$), and two distinguished elements 0 and 1 of $\mathbb{F}$, which satisfy the fundamental properties:

  1. $\mathbb{F}$ is closed under the four operations:

  $$\begin{array}{rcll} + & : & \mathbb{F} \times \mathbb{F} \to \mathbb{F} & \\ * & : & \mathbb{F} \times \mathbb{F} \to \mathbb{F} & \\ - & : & \mathbb{F} \to \mathbb{F} & \\ {}^{-1} & : & \mathbb{F}' \to \mathbb{F}' & \quad \mathbb{F}' = \{a \in \mathbb{F} | a \neq 0\} \end{array}$$

     Of course, we usually write $a + b, a * b, -a$, and a^{-1} instead of $+(a, b), *(a, b), -(a)$, and ${}^{-1}(a)$.

2. '$+$' and '$*$' are commutative and associative, and satisfy the distributive property. That is, for $a, b, c \in \mathbb{F}$:

$$\begin{aligned} a+b &= b+a \\ a*b &= b*a \\ (a+b)+c &= a+(b+c) \\ (a*b)*c &= a*(b*c) \\ a*(b+c) &= a*b+a*c \end{aligned}$$

3. 0 is the identity element and '$-$' is the inverse for addition. 1 is the identity element and 'a^{-1}' is the inverse for multiplication. That is, for $a \in \mathbb{F}$:

$$\begin{aligned} a+0 &= a \\ a+(-a) &= 0 \\ a*1 &= a \\ a*(a^{-1}) &= 1 \qquad (a \neq 0) \end{aligned}$$

4. Although we won't need it for most of linear algebra, I'll mention that $\mathbb{R}$ and $\mathbb{C}$ are both complete (Cauchy sequences have limits), and $\mathbb{R}$ is fully ordered ($a < b$ or $b < a$ or $a = b$ for all $a, b \in \mathbb{R}$).

5. As needed, we will identify $\mathbb{R}$ as a subfield of $\mathbb{C}$, and we will typically write elements of $\mathbb{C}$ as $a + bi$, where a and b are real and $i^2 = -1$.

6. Although $\mathbb{R}$ and $\mathbb{C}$ are the most frequently used coefficient fields, there are many other fields such as $\mathbb{Q}$, the rationals, and the finite fields $\mathbb{Z}/p\mathbb{Z}$ for p a prime.

- If we leave out the requirement that $a * b = b * a$, we get what are called *skew* fields. An important example of a skew field is $\mathbb{H}$, the quaternions (also called the hamiltonians) which contains additional square roots of -1, $i^2 = j^2 = k^2 = -1$, and $ij = k$, $jk = i$, $ki = j$, but $ij = -ji$, $jk = -kj$, and $ki = -ik$. We typically write quaternions either in the form $a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$, or $\alpha + \beta j$ with $\alpha, \beta \in \mathbb{C}$.

- We are then ready for the definition. A **vector space** $V = (V, \mathbb{F}, +, -, *, \vec{0})$ over a field $\mathbb{F}$ is a set V of elements (called vectors) together with a distinguished element $\vec{0}$, two binary operations, and a unary operation:

$$\begin{aligned} + &: V \times V \to V \\ - &: V \to V \\ * &: \mathbb{F} \times V \to V \end{aligned}$$

For $u, v, w \in V$, and $a, b \in \mathbb{F}$, these operations satisfy the properties:

$$\begin{aligned} v+w &= w+v \\ (u+v)+w &= u+(v+w) \\ v+\vec{0} &= v \\ v+(-v) &= \vec{0} \\ 1*v &= v \\ a*(u+v) &= (a*u)+(a*v) \\ (a+b)*v &= (a*v)+(b*v) \\ (a*b)*v &= a*(b*v) \end{aligned}$$

The elements of the field $\mathbb{F}$ are called the **scalars**, or *coefficients* of the vector space.

- From the basic properties listed above, we can prove a variety of additional properties, such as:

$$\begin{aligned} 0 * v &= \vec{0} \\ a * \vec{0} &= \vec{0} \\ (-1) * v &= -v \end{aligned}$$

  We can also prove that the additive identity $\vec{0}$ is unique, as is the additive inverse $-v$.

- We will usually simplify the notation, and write av instead of $a * v$. Furthermore, although it is important to distinguish the scalar 0 from the vector $\vec{0}$, we will typically write the vector in the simple form 0.

- If $\mathbb{F} = \mathbb{R}$, we call V a real vector space, and typically write the scalars as a, b, c. If $\mathbb{F} = \mathbb{C}$, we call V a complex vector space, and often write the scalars as α, β, γ.

Exercises: Vector spaces←

1. Using the basic properties listed above, prove the additional properties:

   (a) $0 * v = \vec{0}$

   (b) $a * \vec{0} = \vec{0}$

   (c) $(-1) * v = -v$

   (d) $-(-v) = v$

   (e) For $a \in \mathbb{F}$ and $v \in \mathbf{V}$, $av = \vec{0}$ iff $a = 0$ or $v = \vec{0}$.

2. Prove that the additive identity $\vec{0}$ is unique.

3. Prove that the additive inverse $-v$ is unique.

Examples of vector spaces ←

- The first main example of a real vector space is $\mathbb{R}^n$, the Cartesian product of n copies of the real line. An element of $\mathbb{R}^n$ looks like $(a_1, a_2, \ldots, a_n)$. When we add two vectors, we get

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n)$$

$$= (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

  When we multiply by a scalar $a \in \mathbb{R}$, we get

$$a * (a_1, a_2, \ldots, a_n) = (aa_1, aa_2, \ldots, aa_n).$$

- Similary, we have the complex vector space $\mathbb{C}^n$, with vector addition and scalar multiplication defined the same way.

- Note that we can also think of $\mathbb{C}^n$ as a real vector space, if we restrict the scalars to $\mathbb{R}$.

- Let $\mathbb{F}^\infty = \{(a_0, a_1, \ldots)\}$, with $(a_0, a_1, \ldots) + (b_0, b_1, \ldots)$
  $= (a_0 + b_0, a_1 + b_1, \ldots)$,
  and $a * (a_0, a_1, \ldots) = (a * a_0, a * a_1, \ldots)$.
  This is a vector space over $\mathbb{F}$.

- For a scalar field $\mathbb{F}$, we can define $\mathbb{F}[x]$ to be the set of all polynomials with formal variable x, and coefficients in $\mathbb{F}$. This is a vector space over $\mathbb{F}$, the coefficient field. Each polynomial is a vector. Vector addition is just polynomial adddition, and scalar multiplication just multiplies each coefficient in the polynomial by the scalar:

$$a * (a_0 + a_1x^1 + \ldots a_nx^n)$$

$$= aa_0 + aa_1x^1 + \ldots aa_nx^n.$$

- For a scalar field $\mathbb{F}$, let $\mathbb{F}^n[x]$ be the set of all polynomials of degree $\leq n$ with coefficients in $\mathbb{F}$. We can also let $\mathbb{F}^\infty[x]$ be the vector space of (formal) power series over $\mathbb{F}$. These are also vector spaces over $\mathbb{F}$.

- Let $C^0(\mathbb{R})$ be the set of all continuous functions with domain and range the real numbers. That is: $C^0(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$. We can define an addition operation on this set. To specify the sum of two functions, we must specify the sum at each point in the domain. If $f, g \in C^0(\mathbb{R})$, we define $f + g$ for each x by

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

  We define scalar multiplication pointwise also: $(a * f)(x) = a * (f(x))$.

  $C^0(\mathbb{R})$ thus becomes a real vector space, where each continuous function is a vector in the space.

- Let $C^n(\mathbb{R})$ be the set of all continuous real functions whose first n derivatives are also continuous, and define addition and scalar multiplication pointwise. Similarly, we can define $C^\infty(\mathbb{R})$ as functions with all derivatives continuous. Again we get real vector spaces.

- Let $C^0[0, 1]$ be the set of all continuous functions with domain the closed interval $[0, 1]$, and range $\mathbb{R}$. We can also define $C^n[0, 1]$ to be those functions with first n derivatives continuous, and $C^\infty[0, 1]$ with all derivatives continuous. We can also generalize to subsets of $\mathbb{R}$ other than the interval $[0, 1]$. With pointwise addition and scalar multiplication, these are each real vector spaces.

- We get can get similar complex vector spaces if we use $\mathbb{C}$ instead of $\mathbb{R}$ in examples like those above.

Exercises: Examples of vector spaces ←

1. Show that each of the examples listed in this section is a vector space.

2. Consider the set of points in $\mathbb{R}^2$ of the forms $(x, 0)$, $x \in \mathbb{R}$, and $(0, y)$, $y \in \mathbb{R}$ (i.e., the union of the X-axis and the Y-axis). Show that with the usual vector addition in $\mathbb{R}^2$, this is not a vector space over $\mathbb{R}$.

3. Consider the set $\mathbb{R}^2$ with usual vector addition:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2),$$

   but with "scalar multiplication" given by

$$a * (x, y) = (ax, y)$$

   for $a \in \mathbb{R}$. Show that this is not a vector space over $\mathbb{R}$.

Subspaces ←

- If V is a vector space over $\mathbb{F}$, a subset $U \subset V$ is called a vector **subspace** of V if U is a vector space over $\mathbb{F}$ using the same vector addition and scalar multiplication as in V. Often, in context, we will just call U a subspace of V.

- Most often, given a subset $U \subset V$, we will just let U inherit the vector addition and scalar multiplication operations from V. Hence, most of the vector space properties will come for free. What may not come for free, however, is whether the subset U is closed under the operations.

Thus, when we inherit the operations from V, we will have that

$$+ : U \times U \to V,$$

and

$$* : \mathbb{F} \times U \to V,$$

whereas, what we need is

$$+ : U \times U \to U,$$

and

$$* : \mathbb{F} \times U \to U.$$

In order for U to be a vector subspace, we must be sure that doing the operations will leave us in the subset U.

- Examples of subspaces:

1. $\{\vec{0}\}$ is a subspace of V for any vector space V.

2. If $m \le n$, then $\mathbb{F}^m$ (identified with $\{(a_1, a_2, \ldots, a_m, 0, \ldots, 0)\}$) is a subspace of $\mathbb{F}^n$.

3. If $m \le n$, then $\mathbb{F}^m[x]$ is a subspace of $\mathbb{F}^n[x]$, and a subspace of $\mathbb{F}[x]$.

4. If $m \le n$, then $C^n(\mathbb{R})$ and $C^\infty(\mathbb{R})$ are subspaces of $C^m(\mathbb{R})$.

5. In $\mathbb{R}^2$, let $U = \{(x, y) \mid y = 3x\}$. Then U is a subspace.

6. More generally, in $\mathbb{R}^n$, fix $a_1, a_2, \ldots, a_n$, and let $U = \{(x_1, x_2, \ldots, x_n) \mid a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0\}$. Then U is a subspace.

7. If U_1 and U_2 are subspaces, so is $U_1 \cap U_2$.

- Examples which are not subspaces:

1. In $\mathbb{R}^2$, let $U = \{(x, y) \mid x = 0 \ or \ y = 0\}$ (i.e., U is the union of the x-axis and the y-axis). Then U is not a subspace, since the two vectors $(1, 0)$ and $(0, 1)$ are in U, but $(1, 0) + (0, 1) = (1, 1)$ is not in U.

2. In $\mathbb{R}^2$, let $U = \{(x, y) \mid x$ and y are both rational numbers$\}$. Then U is not a subspace, since $(1, 1)$ is an element of U, but $\sqrt{2} * (1, 1) = (\sqrt{2}, \sqrt{2})$ is not an element of U.

3. In $\mathbb{R}^n$, fix $a_1, a_2, \ldots, a_n$, and let $U = \{(x_1, x_2, \ldots, x_n) \mid a_1x_1 + a_2x_2 + \ldots + a_nx_n = 2\}$. Then U is not a subspace, since, for example, $\vec{0}$ is not an element of U.

4. In general, if U_1 and U_2 are subspaces, $U_1 \cup U_2$ is not a subspace (except in special cases, such as when $U_1 \subset U_2$).

- If U_1 and U_2 are both subspaces of a vector space V, we can define the subspace which is the **sum** of the two subspaces by:

$U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1$ and $u_2 \in U_2\}$.

- If in addition $U_1 \cap U_2 = \{0\}$, then each element $u \in U_1 + U_2$ can be written in a unique way as $u = u_1 + u_2$. In this case, we call the sum of U_1 and U_2 a **direct sum**, and write it $U_1 \oplus U_2$.

- These ideas generalize in a straightforward way to $U_1 + \cdots + U_n$ and $U_1 \oplus \cdots \oplus U_n$ for a finite number of subspaces, and $U_1 + U_2 + \cdots$ and $U_1 \oplus U_2 \oplus \cdots$ for countably many subspaces.

Exercises: Subspaces ←

1. Show that each of the examples identified in this section as a subspace actually is a vector subspace.

2. Which of the following are vector subspaces of $\mathbb{R}^3$?

   (a) $\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 3x_1 + 2x_2 - x_3 = 0\}$

   (b) $\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 3x_1 + 2x_2 - x_3 = 4\}$

   (c) $\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1x_2x_3 = 0\}$

3. Suppose that U is a vector subspace of V, and V is a vector subspace of W. Show that U is a vector subspace of W.

4. Show that the intersection of any collection of vector subspaces of V is a vector subspace of V.

5. Let
$$l^2 = \{(a_i) \in \mathbb{R}^\infty \mid \sum_{i=0}^{\infty} |a_i|^2 < \infty\}.$$
Show that l^2 is a vector subspace of $\mathbb{R}^\infty$.

6. Let
$$L^2 = \{f \in C^0(\mathbb{R}) \mid \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty\}.$$
Show that L^2 is a vector subspace of $C^0(\mathbb{R})$.

Linear dependence and independence ←

From here on, we'll assume that U and V are vector spaces over a field $\mathbb{F}$.

- Suppose that $v_1, v_2, \ldots, v_n$ are vectors in V, and that for some $a_2, a_3, \ldots, a_n \in \mathbb{F}$,
$$v_1 = a_2v_2 + a_3v_3 + \ldots + a_nv_n.$$
Then we say that v_1 is linearly dependent on $v_2, \ldots, v_n$. Note that if we move v_1 to the other side, and let $a_1 = -1$, we have $a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0$, and not all of the a_i are zero (in particular, $a_1 \neq 0$).

  This motivates a general definition: a set S of vectors in a vector space V is called **linearly dependent** if, for some $n > 0$, and distinct $v_1, v_2, \ldots v_n \in S$, there exist $a_1, a_2, \ldots, a_n \in \mathbb{F}$, not all 0, with
$$a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0.$$

- The flip side of this definition is the following: a set S of vectors in a vector space V is called **linearly independent** if, given distinct $v_1, v_2, \ldots v_n \in S$, the only way for $a_1, a_2, \ldots, a_n \in \mathbb{F}$ to give
$$a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0$$
is if every one of the a_i are zero.

  A useful way to think about this is that a set S of vectors is linearly independent if no individual one of the vectors is linearly dependent on a finite number of the rest of them.

- Some examples:

  1. In $\mathbb{R}^2$, the sets of vectors $\{(1,0),(0,1)\}$, $\{(1,2),(1,3)\}$, $\{(1,0)\}$ and $\{(2,2),(3,2)\}$ are each linearly independent sets.

2. In $\mathbb{R}^2$, none of the sets of vectors $\{(1,0),(2,0)\}$, $\{(1,0),(0,1),(1,1)\}$, $\{(0,0),(0,1)\}$, $\{(0,0)\}$, $\{(1,2),(2,3),(3,4),(4,1)\}$, nor $\{(1,2),(3,4),(2,1)\}$ is a linearly independent set.

3. In any vector space, if a set of vectors contains the 0 vector, the set is not linearly independent.

4. In $C^0(\mathbb{R})$, if $f_1(1) = f_2(2) = \ldots = f_n(n) = 1$, but $f_i(j) = 0$ for $i \neq j$, then the set of functions $\{f_1, f_2, \ldots, f_n\}$ is linearly independent.

5. In $C^\infty(\mathbb{R})$, the set of functions $\{\cos(nx) \mid n = 0,1,2,\ldots\}$ is a linearly independent set.

6. The empty set, $\{\}$, is a linearly independent set.

- Mathematics is an extremely precise language. These two sentences do not mean the same thing:

  1. The set $\{v_1, v_2, \ldots, v_n\}$ of vectors is linearly independent.

  2. $\{v_1, v_2, \ldots, v_n\}$ is a set of linearly independent vectors.

  One of the hardest parts of doing mathematics is developing your mathematical intuition. It is tempting to imagine that intuition is what you have before you know anything, but that is nonsense. Intuition is just the automatic part of your knowledge, derived from your past experience. Becoming better at mathematics involves learning new mathematics, and then integrating that new knowledge into your intuition. Doing that takes care, precision, and **lots of practice**!

Exercises: Linear dependence and independence ←

1. Verify the statements in each of the six examples in this section.

2. Suppose that $P_0, P_1, \ldots, P_n$ are polynomials in $\mathbb{F}^n[x]$, and $P_i(1) = 0$ for $i = 0,1,\ldots,n$. Show that $\{P_0, P_1, \ldots, P_n\}$ is not linearly independent in $\mathbb{F}^n[x]$.

3. In $C^\infty(\mathbb{R})$, show that each of these sets of functions is linearly independent:

   (a) $\{x^n \mid n = 0,1,2,\ldots\}$.

   (b) $\{\sin(nx) \mid n = 0,1,2,\ldots\}$.

   (c) $\{e^{nx} \mid n = 0,1,2,\ldots\}$.

Span of a set of vectors ←

- If S is a set of vectors in the vector space V over $\mathbb{F}$, we define the **span** of the set S by:

$$\text{span}(S) = \{a_1v_1 + a_2v_2 + \ldots + a_nv_n \mid a_i \in \mathbb{F}, v_i \in S, n > 0\}.$$

  This says that span(S) is the set of all vectors that can be written as finite linear combinations of the vectors in S.

- Examples:

  1. In $\mathbb{R}^1$, span($\{(1)\}$) is all of $\mathbb{R}^1$.

  2. In $\mathbb{R}^3$, span($\{(1,0,0),(0,1,0)\}$) is the $x-y$ plane. Similarly, span($\{(1,0,0),(0,0,1)\}$) is the $x-z$ plane, span($\{(0,1,0),(0,0,1)\}$) is the $y-z$ plane, and span($\{(1,0,0),(0,1,0),(0,0,1)\}$) is all of $\mathbb{R}^3$.

3. Still in $\mathbb{R}^3$, span$(\{(1,2,0),(2,1,0)\})$ is the $x-y$ plane, span$(\{(1,0,0)\})$ is the x axis,
span$(\{(1,0,0),(1,1,0),(1,1,1)\})$ is all of $\mathbb{R}^3$,
span$(\{(1,2,3),(1,1,0),(1,1,1),(2,1,4)\})$ is all of $\mathbb{R}^3$, and
span$(\{(1,2,0),(1,1,0),(3,1,0),(2,1,0)\})$ is the $x-y$ plane.

4. In $\mathbb{F}^n[x]$, span$(\{1,x^1,x^2,\ldots,x^n\}) = \mathbb{F}^n[x]$.

5. In $\mathbb{F}[x]$, span$(\{1,x^1,x^2,\ldots\}) = \mathbb{F}[x]$.

6. In $C^0(\mathbb{R})$, span$(\{f_1(x) = 1, f_2(x) = x\})$ is the set of all linear functions, $y = mx + b$.

7. For any set S, span(S) is a subspace.

8. In particular, span(S) is sometimes defined to be the smallest subspace of V containing S, or the intersection of all subspaces of V that contain S.

Exercises: Span of a set of vectors ←

1. Verify the statements in each of the eight examples in this section.

2. Show that if span$(\{v_0,v_1,\ldots,v_n\}) = V$, then span$(\{v_0 - v_1, v_1 - v_2, \ldots, v_{n-1} - v_n, v_n\}) = V$.

Basis for a vector space ←

- Suppose S is an ordered set (or *list*) of vectors in a vector space V. Suppose further that:

  1. S is linearly independent, and

  2. span$(S) = V$.

  Then we call S a **basis** for the vector space V.

- Given a basis S for the vector space V, every vector $v \in V$, $v \neq 0$, can be written in a unique way as:

$$v = a_1v_1 + a_2v_2 + \ldots + a_nv_n$$

  with $a_1, a_2, \ldots, a_n \in \mathbb{F}$, $a_i \neq 0$, and $v_1, v_2, \ldots, v_n \in S$. For uniqueness, we take the v_i to be distinct, and to be in the order they appear in S. We are writing each v as a finite linear combination of the basis vectors.

- Examples:

  1. In $\mathbb{F}^3$, the ordered set $S = ((1,0,0),(0,1,0),(0,0,1))$ is a basis (often called the **standard basis**). This generalizes in the obvious way to $\mathbb{F}^n$.

  2. In $\mathbb{F}^3$, the ordered set $S = ((1,0,0),(1,1,0),(1,1,1))$ is a basis. So is $((1,2,3),(2,1,3),(1,2,2))$.

  3. In general, in $\mathbb{F}^n$ any linearly independent ordered set $S = (v_1, v_2, \ldots, v_n)$ of size n is a basis.

  4. $(1, x, x^2, x^3, \ldots)$ is a basis for $\mathbb{F}[x]$.

  5. If S is a linearly independent ordered set, then S is a basis for span(S).

- If a vector space V has a finite basis $S = (v_1, v_2, \ldots, v_n)$, we say that V is finite dimensional, and we define the **dimension** of V by:

$$\dim(V) = n.$$

We define $\dim(\{\vec{0}\}) = 0$.

Note that if a vector space V has a finite basis of size n, then every basis for V contains n vectors, and thus the definition makes sense.

For example, $\dim(\mathbb{F}^n) = n$ for any field $\mathbb{F}$ and $n > 0$.

We also have that $\dim(\mathbb{F}^n[x]) = n + 1$.

- If a vector space V has a finite or countably infinite basis S, then we can uniquely represent each vector v with respect to S by a list of the form:
$v = (b_1, b_2, \ldots, b_n)$ in the finite case, or
$v = (b_1, b_2, \ldots)$ in the infinite case.

Each b_i is either the non-zero coefficient corresponding with the ith element of S from the unique representation described above, or 0 if the basis element does not appear there. In the infinite case, only finitely many of the b_i are non-zero.

We represent 0 by $(0, 0, \ldots, 0)$ in the finite case, and by $(0, 0, \ldots)$ in the infinite case.

- In these cases, we also have, for $v_i \in S$, that

$$V = \text{span}(v_1) \oplus \cdots \oplus \text{span}(v_n)$$

in the finite case, and

$$V = \text{span}(v_1) \oplus \text{span}(v_2) \oplus \cdots$$

in the countably infinite case.

Exercises: Basis for a vector space ←

1. Verify the statements in each of the five examples in this section.

2. Let U be the subspace of $\mathbb{R}^6$ given by $U = \{(x_1, x_2, x_3, x_4, x_5, x_6) \in \mathbb{R}^6 \mid 2x_1 + 3x_3 = 0 \text{ and } x_2 = 4x_5\}$.

   Find a basis for U.

3. Show that if U_1 and U_2 are subspaces of a finite dimensional vector space, then $\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2)$.

Linear transformations ←

- If U and V are vector spaces over $\mathbb{F}$, then a function $T : U \to V$ is called a **linear transformation** or **linear mapping** if

$$T(a_1u_1 + a_2u_2) = a_1T(u_1) + a_2T(u_2)$$

for all $a_1, a_2 \in \mathbb{F}$ and $u_1, u_2 \in U$.

An equivalent pair of conditions is that $T(u_1 + u_2) = T(u_1) + T(u_2)$, and $T(au) = aT(u)$.

- For a linear transformation $T : U \to V$, we call U the **domain** and V the **codomain** (or sometimes **range**) of T. We also define the **kernel** (or **null space**) of T by:

$$\ker(T) = \{u \in U \mid T(u) = 0\}.$$

Further, we define the **image** (or sometimes **range** – be careful here!) of T by:

$$\text{im}(T) = \{v \in V \mid v = T(u) \text{ for some } u \in U\}.$$

- For a linear transformation $T : U \to V$ we have the nice properties:

  1. ker(T) is a subspace of U, and im(T) is a subspace of V. (ex)

  2. If U is finite dimensional, then
  $$\dim(U) = \dim(\ker(T)) + \dim(\operatorname{im}(T)).$$
  (ex)

  3. If U is finite dimensional with basis $(u_1, u_2, \ldots, u_n)$, and V is finite dimensional with basis $(v_1, v_2, \ldots, v_m)$, then T is determined by its effect on the basis elements u_i. There exist a_{ij}, $1 \le i \le m$, $1 \le j \le n$ with:
  $$T(u_j) = a_{1j}v_1 + a_{2j}v_2 + \ldots + a_{mj}v_m,$$
  and in general, if $u = \sum_j b_j u_j$, then
  $$T(u) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_j v_i.$$

4. Thus, given particular bases for U and V, we can represent the linear transformation T by the matrix
$$[T] = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$
or $[T]_{ij} = a_{ij}$.

5. If we represent the vector $u \in U$ by the column matrix $[b_1, b_2, \ldots, b_n]^t$, then we have
$$\begin{aligned} [T(u)] &= [T][u] \\ &= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + \ldots + a_{1n}b_n \\ a_{21}b_1 + a_{22}b_2 + \ldots + a_{2n}b_n \\ \vdots \\ a_{m1}b_1 + a_{m2}b_2 + \ldots + a_{mn}b_n \end{bmatrix} \end{aligned}$$

6. Sometimes it is more convenient to use the Einstein summation conventions, where we would write:
$$T(u)_i = T_i^j u_j$$
with implied summation over the repeated upper and lower indices.

7. If $T_1 : U \to V$ and $T_2 : U \to V$ are two linear transformations, we can define the sum of the two transformations as
$$(T_1 + T_2)(u) = T_1(u) + T_2(u),$$
and we can define scalar multiplication of a linear transformation T by a as
$$(a * T)(u) = a * (T(u)).$$
We can thus define $L(U, V)$ to be the space of all linear transformations from U to V. We define the **zero** transformation $0 : U \to V$ by $0(u) = \vec{0}$. With these definitions, $L(U, V)$ is a vector space. (ex)

8. Given particular bases for U and V, the matrix representation of $T_1 + T_2$ is given by: $[T_1 + T_2] = [T_1] + [T_2]$.

9. If U and V are finite dimensional, then $L(U, V)$ is also finite dimensional, and
$$\dim(L(U, V)) = \dim(U) * \dim(V).$$
(ex)

10. If $T : U \to V$ and $S : V \to W$ are linear transformations, then the composition $S \circ T : U \to W$ is also a linear transformation. (ex)

    Recall that $(S \circ T)(u) = S(T(u))$.

    Given particular bases for U, V, and W, the matrix representation of $S \circ T$ is the matrix product $[S \circ T] = [S][T]$. We usually abbreviate $S \circ T$ as ST.

    It is worth noting that unless $W \subseteq U$, it doesn't even make sense to talk about $T \circ S$.

11. The matrix of the composition of two linear transformations is given by the product of the two matrices, given by:

$$[ST] = [S][T] =$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} \sum_{i=1}^n a_{1i}b_{i1} & \sum_{i=1}^n a_{1i}b_{i2} & \dots & \sum_{i=1}^n a_{1i}b_{ip} \\ \sum_{i=1}^n a_{2i}b_{i1} & \sum_{i=1}^n a_{2i}b_{i2} & \dots & \sum_{i=1}^n a_{2i}b_{ip} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n a_{mi}b_{i1} & \sum_{i=1}^n a_{mi}b_{i2} & \dots & \sum_{i=1}^n a_{mi}b_{ip} \end{bmatrix}$$

12. We also have the distributive property:

$$S(T_1 + T_2) = (ST_1) + (ST_2).$$

The matrix representation for this is:

$$[S][T_1 + T_2] = [S][T_1] + [S][T_2].$$

- Some examples of linear transformations:

1. The function $T : \mathbb{R}^2 \to \mathbb{R}$ given by $T((x, y)) = x + y$ is linear.

2. The function $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by $T((x, y)) = (x + 2y, x - 2y, x - y)$ is linear.

3. Given $a_1, a_2, \dots, a_n \in \mathbb{F}$, and V a finite dimensional vector space over $\mathbb{F}$ with basis $(v_1, v_2, \dots, v_n)$, the function $T : V \to \mathbb{F}$ given by $T(b_1v_1 + \dots + b_nv_n) = a_1b_1 + \dots + a_nb_n$ is linear.

   In general, for a vector space V, a linear transformation $T : V \to \mathbb{F}$ is called a **linear functional**. The study of these transformations is called *functional analysis*.

4. The function $D : \mathbb{F}^n[x] \to \mathbb{F}^{n-1}[x]$ given by
   $D(a_nx^n + \dots + a_1x + a_0)$
   $= na_nx^{n-1} + \dots + 2a_2x + a_1$
   is linear. This linear transformation is called the derivative ...

5. Similarly, we have the derivative transformation D : $\mathbb{F}^\infty[x] \to \mathbb{F}^\infty[x]$ given by

$$\mathrm{D}\Big(\sum_{n=0}^{\infty} a_nx^n\Big) = \sum_{n=1}^{\infty} n * a_nx^{n-1}.$$

6. The shift map $S : \mathbb{F}^\infty \to \mathbb{F}^\infty$ is given by: $S((a_0, a_1, \dots)) = (a_1, a_2, \dots)$.

7. The difference map $\Delta : \mathbb{F}^\infty \to \mathbb{F}^\infty$ is given by:

$$\Delta((a_0, a_1, \dots)) = (a_1 - a_0, a_2 - a_1, \dots),$$

   or, abbreviating $(a_0, a_1, \dots)$ by (a_n),

$$\Delta((a_n)) = (a_{n+1} - a_n).$$

Exercises: Linear transformations ←

1. Verify each of the statements marked (ex) in this section.

2. Verify that each of the 7 examples actually are linear transformations.

3. Show that the usual calculus derivative $\frac{d}{dx} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ given by:

$$\frac{d\ f(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

   is a linear transformation.

4. Is the usual calculus indefinite integral

$$\int f(x)dx$$

   a linear transformation? Why or why not? What about the definite integral?

Morphisms – mono, epi, and iso ←

- In algebra, we call a function that preserves structure a **morphism**. In our current context, a linear transformation preserves the linear structure of a vector space in the sense that $T(u_1 + u_2) = T(u_1) + T(u_2)$ and $T(au) = aT(u)$. Thus, linear transformations are morphisms in the category of vector spaces.

- We will be particularly interested in morphisms that have additional properties. Specifically, we are likely to look for morphisms $T : U \to V$ which have one or more of the properties:

1. One-to-one. Such transformations are also called **injective**, or **monomorphisms**. These transformations have the property that if $u_1 \neq u_2$, then $T(u_1) \neq T(u_2)$. This says that different things get sent different places – that is, no two things get sent to the same place.

   Monomorphisms are nice because the subspace im$(T) \subset V$ looks just like U.

2. Onto. Such transformations are also called **surjective**, or **epimorphisms**. These transformations have the property that for every $v \in V$, there exists some $u \in U$ with $T(u) = v$. This says that every element of V is hit by some element of U. Another way to say this is that im$(T) = V$.

   Epimorphisms are nice, because the algebraic properties of V will be reflected back in U.

3. Both one-to-one and onto. Such transformations are also called **bijective**. A bijective function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ with the properties that for all $x \in X$ and all $y \in Y$, $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$.

   A bijective morphism whose inverse also preserves algebraic structure is called an **isomorphism**. In linear algebra, we have the nice property that if a linear transformation is bijective, then its inverse is also linear, and thus it is an isomorphism.

4. If there is an isomorphism $T : U \to V$, we say that the spaces U and V are **isomorphic**. If two spaces are isomorphic, then they share all relevant properties. Thus, two isomorphic vector spaces are indistinguishable as vector spaces except for a renaming of the elements.

- We have the nice fact that a linear transformation $T : U \to V$ is a monomorphism (is one-to-one) if and only if ker$(T) = \{0\}$.

  Pf.: Suppose T is a monomorphism. We know that for every linear transformation, $T(0) = 0$. Then, since T is a monomorphism, we know that if $T(u) = 0 = T(0)$, it must be that $u = 0$. Thus ker$(T) = \{0\}$.

  On the other hand, suppose that ker$(T) = \{0\}$. Then, if $T(u_1) = T(u_2)$, we will have $0 = T(u_1) - T(u_2) = T(u_1 - u_2)$. But this means that $u_1 - u_2 \in$ ker(T) and hence, since we are assuming that ker$(T) = \{0\}$, we must have $u_1 - u_2 = 0$, or $u_1 = u_2$. By the contrapositive, this means that if $u_1 \neq u_2$, then $T(u_1) \neq T(u_2)$. Q.E.D.

  (I had to do at least one proof, didn't I? :-)

- An important example of an isomorphism is the identity transformation $I_U : U \to U$ given by $I_U(u) = u$. (ex)

- If $T : U \to V$ is a monomorphism, then there is a linear transformation $S_1 : \mathrm{im}(T) \to U$ with $S_1(T(u)) = (S_1T)(u) = u$ for all $u \in U$. This says that $S_1T = I_U$. S_1 is called a left (partial) inverse of T. (ex)

- If $T : U \to V$ is an epimorphism, then there is a linear transformation $S_2 : V \to U$ with $T(S_2(v)) = (TS_2)(v) = v$ for all $v \in V$. This says that $TS_2 = I_V$. S_2 is called a right (partial) inverse of T. (ex)

- If $T : U \to V$ is an isomorphism, then since T is an epimorphism, both S_1 and S_2 (as above) exist. Also, $\mathrm{im}(T) = V$. We therefore have that for all $v \in V$,

  $$S_1(v) = S_1((TS_2)(v)) = (S_1T)(S_2(v)) = S_2(v).$$

  Thus $S_1 = S_2$. In this case, the (two-sided) inverse of T exists, and we have $T^{-1} = S_1 = S_2$.

- If U and V are finite dimensional vector spaces over $\mathbb{F}$ with $\dim(U) = \dim(V)$, then U and V are isomorphic. (ex)

  (Big) hint for proof: Let $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ be bases for U and V respectively. Define $T : U \to V$ by $T(u_i) = v_i$ for $1 \le i \le n$, and extend by linearity. Make sense of the phrase "extend by linearity," and then show that T is an isomorphism.

- We also have the nice fact that if $\dim(U) = \dim(V) = n,\ U,\ V$ over $\mathbb{F}$, and $T : U \to V$ is linear, then the following are equivalent: (ex)

  1. T is a monomorphism.

  2. T is an epimorphism.

  3. T is an isomorphism.

  This means, for example, that such a T is onto if and only $\ker(T) = 0$.

- If $T : U \to V$ and $S : V \to W$, then (ex)

  1. If both S and T are monomorphisms, then so is ST.

  2. If ST is a monomorphism, then so is T.

  3. If both S and T are epimorphisms, then so is ST.

  4. If ST is an epimorphism, then so is S.

- If $T : V \to V$ and $S : V \to V$, and $ST = TS$, then (ex)

  1. Both S and T are monomorphisms if and only if ST is a monomorphism.

  2. Both S and T are epimorphisms if and only if ST is an epimorphism.

- Let $\mathrm{matrix}(\mathbb{F}, n, m)$ be the space of all n by m matrices with entries from $\mathbb{F}$. We use ordinary entry by entry matrix addition, where $(A + B)_{ij} = A_{ij} + B_{ij}$, scalar multiplication, where $(aA)_{ij} = aA_{ij}$, and let $(0)_{ij} = 0$. Then $\mathrm{matrix}(\mathbb{F}, n, m)$ is an $n * m$-dimensional vector space over $\mathbb{F}$. If $\dim(U) = m$ and $\dim(V) = n$, then we can define

  $$\mathrm{mat} : L(U, V) \to \mathrm{matrix}(\mathbb{F}, n, m)$$

  by $\mathrm{mat}(T) = [T]$.

  This transformation is an isomorphism. (ex)

Exercises: Morphisms – mono, epi, and iso ←

1. Verify each of the statements marked (ex) in this section.

2. Show that if a function has an inverse, then the inverse is unique.

3. Consider the function $T : \mathbb{F}^n[x] \to \mathbb{F}^{n+1}$ given by
$$T(a_0 + a_1 x + \ldots + a_n x^n) = (a_0, \ldots, a_n).$$
Show that this function is an isomorphism.

4. Consider the function $T : \mathbb{C} \to \text{matrix}(\mathbb{R}, 2, 2)$ given by
$$T(a+bi) \quad = \quad \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

(a) Show that if we consider $\mathbb{C}$ as a real vector space, this function is a monomorphism.

(b) Show that this function also respects complex multiplication, that is,
$$T((a+bi)(c+di)) = T(a+bi)T(c+di).$$

5. More generally, consider the function $T : \text{matrix}(\mathbb{C}, n, n) \to \text{matrix}(\mathbb{R}, 2n, 2n)$ given by
$$T\left(\begin{bmatrix} a_{11}+b_{11}i & \cdots & a_{1n}+b_{1n}i \\ \vdots & & \vdots \\ a_{n1}+b_{n1}i & \cdots & a_{nn}+b_{nn}i \end{bmatrix}\right)$$
$$= \begin{bmatrix} a_{11} & -b_{11} & \cdots & a_{1n} & -b_{1n} \\ b_{11} & a_{11} & \cdots & b_{1n} & a_{1n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & -b_{n1} & \cdots & a_{nn} & -b_{nn} \\ b_{n1} & a_{n1} & \cdots & b_{nn} & a_{nn} \end{bmatrix}$$

(a) Show that if we consider $\text{matrix}(\mathbb{C}, n, n)$ as a real vector space, this function is a monomorphism.

(b) Show that this function also respects matrix multiplication, that is,
$$T(AB) = T(A)T(B).$$

6. What the heck. Let $\mathbb{H}$ be the quaternions (described above). Consider the function $T : \text{matrix}(\mathbb{H}, n, n) \to \text{matrix}(\mathbb{C}, 2n, 2n)$ given by
$$T\left(\begin{bmatrix} \alpha_{11}+\beta_{11}j & \cdots & \alpha_{1n}+\beta_{1n}j \\ \vdots & & \vdots \\ \alpha_{n1}+\beta_{n1}j & \cdots & \alpha_{nn}+\beta_{nn}j \end{bmatrix}\right)$$
$$= \begin{bmatrix} \alpha_{11} & -\overline{\beta_{11}} & \cdots & \alpha_{1n} & -\overline{\beta_{1n}} \\ \beta_{11} & \alpha_{11} & \cdots & \beta_{1n} & \alpha_{1n} \\ \vdots & \vdots & & \vdots & \vdots \\ \alpha_{n1} & -\overline{\beta_{n1}} & \cdots & \alpha_{nn} & -\overline{\beta_{nn}} \\ \beta_{n1} & \alpha_{n1} & \cdots & \beta_{nn} & \alpha_{nn} \end{bmatrix}$$

(a) Show that if we consider $\text{matrix}(\mathbb{H}, n, n)$ as a complex vector space, this function is a monomorphism.

(b) Show that this function also respects matrix multiplication, that is,
$$T(AB) = T(A)T(B).$$

Thus, if we denote by $\mathbf{O}(n)$, $\mathbf{U}(n)$, and $\mathbf{Sp}(n)$ the distance preserving linear operators on $\mathbb{R}^n$, $\mathbb{C}^n$, and $\mathbb{H}^n$ respectively (called the orthogonal, unitary, and symplectic groups), then we have the monomorphisms:
$$\cdots \to \mathbf{O}(n) \to \mathbf{Sp}(n) \to \mathbf{U}(2n) \to \mathbf{O}(4n)$$
$$\to \mathbf{Sp}(4n) \to \mathbf{U}(8n) \to \mathbf{O}(16n) \to \cdots$$

Linear operators $\leftarrow$

- If $T : V \to V$ is a linear transformation, we call T a **linear operator** on V. Note that if S and T are operators on V, then so is ST. We can abbreviate $L(V, V)$ as $L(V)$.

  $L(V)$ has the algebraic structure of a **ring** with identity. A ring is similar to a field (as defined above), except without the requirements that multiplication be commutative and that there be multiplicative inverses for non-zero elements. The identity element is I_V. $L(V)$ is a non-commutative ring, since in general $ST \neq TS$. This is reflected in the fact that matrix multiplication is non-commutative. Only in very special cases is it true that $[S][T] = [T][S]$ (for example, if both $[S]$ and $[T]$ are diagonal matrices, with $[S]_{ij} = 0$ for $i \neq j$, and $[T]_{ij} = 0$ for $i \neq j$).

- Note that for operators, it makes sense to talk about $T \circ T$, and we can thus define $T^n = T \circ T \circ \ldots \circ T$ (n times). We also define $T^0 = I$, the identity operator.

- Thus, if we have a polynomial $P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ from $\mathbb{F}[x]$, we can talk about the polynomial in T:

  $$P(T) = a_n T^n + a_{n-1} T^{n-1} + \ldots + a_1 T + a_0.$$

  (The a_0 term stands for $a_0 I$.) This is an operator on V which acts on vectors as:

  $$P(T)(v) = a_n T^n(v) + \ldots + a_1 T(v) + a_0 v.$$

- In fact, we can generalize this to power series. If

  $$p(x) = \sum_{n=0}^{\infty} a_n x^n,$$

  then we can define

  $$p(T) = \sum_{n=0}^{\infty} a_n T^n.$$

- This means, for example, that we can define the exponential of an operator:

  $$\exp(T) = \sum_{n=0}^{\infty} \frac{1}{n!} T^n.$$

- We can even define the cosine or sine of an operator, etc.:

  $$\cos(T) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} T^{2n},$$

  $$\sin(T) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} T^{2n+1}.$$

- We can talk about square roots of an operator, saying that S is a square root of T if $S^2 = T$. It is unlikely that S is unique.

- We can also talk about logarithms of an operator, saying that S is a logarithm of T if $\exp(S) = T$. Again, it is unlikely that S is unique.

- Examples of linear operators:

  1. A first important example begins like this: suppose we have a system of m linear equations in n variables

     $$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= 2_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m. \end{aligned}$$

     If we let $A : \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation with $[A]_{ij} = a_{ij}$, $\mathbf{x}$ be the vector $(x_1, x_2, \ldots, x_n)^t$, and $\mathbf{b}$ the vector $(b_1, b_2, \ldots, b_m)^t$, then we can rewrite the equation in the form:

     $$A\mathbf{x} = \mathbf{b}.$$

     We then have several possibilities.

(a) The first case is when $m = n$. There are then two possibilities:

i. If A is an isomorphism, then A^{-1} exists, and we can solve the equation for $\mathbf{x}$:

$$\mathbf{x} = A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}.$$

ii. If A is not an isomorphism, then in particular A is not a an epimorphism, and thus it is possible that $\mathbf{b} \notin \text{im}(A)$. In this case, there are no solutions to the equation.

On the other hand, if $\mathbf{b} \in \text{im}(A)$ there is at least one solution $\mathbf{x}_0$ with $A\mathbf{x}_0 = \mathbf{b}$.

Note, though, that we also know A is not a monomorphism, and hence $\dim(\ker(A)) \geq 1$. Then, if $\mathbf{z} \in \ker(A)$, we have $A(\mathbf{x}_0 + \mathbf{z}) = A\mathbf{x}_0 + A\mathbf{z} = A\mathbf{x}_0 + 0 = A\mathbf{x}_0 = \mathbf{b}$, and so $\mathbf{x}_0 + \mathbf{z}$ is another solution.

Furthermore, if $\mathbf{y}$ is another solution with $A\mathbf{y} = \mathbf{b}$, then $A(\mathbf{x}_0 - \mathbf{y}) = A\mathbf{x}_0 - A\mathbf{y} = \mathbf{b} - \mathbf{b} = 0$. This means $\mathbf{x}_0 - \mathbf{y} \in \ker(A)$, and so $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ for some $\mathbf{z} \in \ker(A)$.

Thus, if $\mathbf{x}_0$ is a particular solution, then every solution is of the form $\mathbf{x}_0 + \mathbf{z}$ for some $\mathbf{z} \in \ker(A)$. The space of solutions is then a translation of the kernel of A, of the form $\mathbf{x}_0 + \ker(A)$. We then only need to find one particular solution.

In this case, we have broken the problem down into two parts: first, we solve $A\mathbf{x} = 0$ (called the *homogeneous* equation), then we find a single solution $A\mathbf{x}_0 = \mathbf{b}$. For $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, there will be infinitely many solutions.

(b) The second case is when $m < n$. In this case, A cannot be a monomorphism, and things are then similar to the second part of the previous case. If A is an epimorphism, there is sure to be at least one solution. Again, we solve the homogeneous case, and then find one particular solution. If A is not an epimorphism, there may not be any solutions. Otherwise, as above, in general we will have infinitely many solutions.

(c) If $m > n$, then A cannot be an epimorphism, and hence there may not be any solutions. If A is a monomorphism, if there is a solution, there will be just one. If A is not a monomorphism, if there is one solution, there will be infinitely many, as above.

(d) In all three cases, we can start by trying to find the inverse (or left partial inverse) of A. If A is a monomorphism, we can expect to be successful, and to be able to find the unique solution if it exists. If A is not a monomorphism, we solve the homogeneous equation, and then look for a single particular solution.

(e) Question: How can we tell if A is a monomorphism? How can we find A^{-1}, or at least S_1, the left partial inverse of A?

2. A second important example is the derivative operator

$$\mathsf{D} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}).$$

Note that this is a linear operator, since $\mathsf{D}(f+g) = \mathsf{D}(f) + \mathsf{D}(g)$, and $\mathsf{D}(af) = a\mathsf{D}(f)$.

Note that ker(D) = span({1}), the one-dimensional space consisting of all constant functions. If we collapse ker(D) down to nothing (in technical terms, form the *quotient space* ...) then we can think of D as an isomorphism (on the quotient space). D has an inverse,

$$\int : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}).$$

We then have (a variant of) the **fundamental theorem of calculus**:

$$\mathsf{D}\int(f) = \int \mathsf{D}(f) = f$$

(on the reduced space – that is, up to a constant).

3. We can then consider the general linear ordinary differential operators with constant coefficients. These are operators in $\mathbb{R}[\mathsf{D}]$, that is operators of the form:

$$P(\mathsf{D}) = a_n\mathsf{D}^n + \ldots + a_1\mathsf{D} + a_0.$$

These operators act on a function $f \in C^\infty(\mathbb{R})$ as

$$P(\mathsf{D})(f) = a_n\mathsf{D}^n(f) + \ldots + a_1\mathsf{D}(f) + a_0 f.$$

We then have the general nth order linear ordinary differential equation with constant coefficients:

$$P(\mathsf{D})(f) = g.$$

We can work on solving these equations with an approach similar to the method for systems of linear equations.

We first note that if the function f_0 is a solution to this equation, and $z = z(x) \in \ker(P(\mathsf{D}))$, then $f_0 + z$ is also a solution, and if f_1 is another solution, then $P(\mathsf{D})(f_0 - f_1) = P(\mathsf{D})(f_0) - P(\mathsf{D})(f_1) = g - g = 0$. Thus, all solutions are of the form $f_0 + z$ where f_0 is some particular solution, and $z \in \ker(P(\mathsf{D}))$.

We thus separate the problem into two parts. First we solve the associated *homogeneous* equation:

$$P(\mathsf{D})(f) = 0$$

to find $\ker(P(\mathsf{D}))$, and then we look for a single particular solution to the original equation.

In general, we have that $\dim(\ker(P(\mathsf{D}))) = n$, the degree of P. This is not an entirely obvious fact, but it is not counterintuitive ...

Hence what we need to do is find n functions which form a basis for $\ker(P(\mathsf{D}))$. What we need, then, are n linearly independent functions each of which is a solution to the homogeneous equation.

In theory (:-) this is not too hard.

We note first that for the first-order case, we have the solution:

$$\begin{aligned}(\mathsf{D} - r)(e^{rx}) &= \mathsf{D}(e^{rx}) - re^{rx} \\ &= re^{rx} - re^{rx} \\ &= 0,\end{aligned}$$

and, in the kth order extension of this,

$$(\mathsf{D} - r)^k(x^{k-1}e^{rx}) = 0.$$

We also have that the set of functions $A = \{x^j e^{r_i x} \mid 0 \le j \le k, 1 \le i \le m\}$ is linearly independent. From this we see how to solve equations of the form:

$$\prod_i (\mathsf{D} - r_i)^{k_i}(f) = 0.$$

Now, consider the operator

$$\mathsf{D}^2 - 2s\mathsf{D} + s^2 + t^2$$

for $s, t \in \mathsf{R}$.

We have that

$$(\mathsf{D}^2 - 2s\mathsf{D} + s^2 + t^2)^k (x^{k-1} e^{sx} \cos(tx)) = 0,$$

$$(\mathsf{D}^2 - 2s\mathsf{D} + s^2 + t^2)^k (x^{k-1} e^{sx} \sin(tx)) = 0.$$

We can think of this as $\alpha = s + ti$, and that we are working with

$$\mathsf{D}^2 - 2s\mathsf{D} + s^2 + t^2 = (\mathsf{D} - \alpha)(\mathsf{D} - \bar{\alpha}).$$

We have that the set of functions

$$B = \{x^j e^{s_k x} \cos(t_k x), x^m e^{s_n x} \sin(t_n x)\}$$

is also linearly independent. From this we see how to solve equations of the form:

$$\prod_i (\mathsf{D}^2 - 2s_i\mathsf{D} + s_i^2 + t_i^2)^{k_i}(f) = 0.$$

We can now put together all the pieces to solve the homogeneous equation $P(\mathsf{D}) = 0$. We use the fact that any polynomial over $\mathbb{R}$ can be completely factored as

$$P(\mathsf{D}) = \prod_i (\mathsf{D} - r_i)^{k_i} \prod_j (\mathsf{D} - \alpha_j)^{k_j} (\mathsf{D} - \bar{\alpha}_j)^{k_j}$$

with $r_i \in \mathsf{R}$ and $\alpha_j \in \mathsf{C}$.

To solve the inhomogeneous equation, we need only to find one particular solution of $P(\mathsf{D})(f) = g$.

This is just the bare beginnings of techniques for solving differential equations, but it gives the flavor of some relatively powerful methods, and the role that linear algebra plays. I haven't even mentioned the issues of initial values/boundary conditions. For much more on these topics, look at a book such as *Elementary Differential Equations with Linear Algebra,* by Finney and Ostberg.

4. These approaches generalize to linear ordinary differential operators with non-constant coefficients:

   $$a_n(x)\mathsf{D}^n + a_{n-1}(x)\mathsf{D}^{n-1} + \cdots + a_0(x),$$

   to systems of linear ordinary differential operators, and on to linear partial differential operators and systems of linear partial differential operators. But I think for now someone else will have to write that up ... :-)

5. We can develop a similar approach to solving linear discrete difference equations. For example, the difference equation $(a_{n+2} - a_{n+1} - a_n) = (0)$ has as a solution the Fibonnacci sequence $(a_n) = (1, 1, 2, 3, 5, 8, \ldots)$. We could work with the discrete version Δ of the differential operator D, where $\Delta f(x) = (f(x+1) - f(x))/1 = f(x+1) - f(x)$. In the discrete case, we can't let h go to zero.

We would then have $\Delta((a_n)) = (a_{n+1} - a_n)$. Our example difference equation would be

$$(\Delta^2 + \Delta - 1)((a_n)) = (0).$$

Often, however, it is more convenient to work with the discrete increment operator $\mathbf{E}$, with $\mathbf{E}(f(x)) = f(x+1)$, or $\mathbf{E}((a_n)) = (a_{n+1})$. Both Δ and $\mathbf{E}$ are linear operators. Our example equation is then $\mathbf{E}^2((a_n)) = \mathbf{E}((a_n)) + (a_n)$, or

$$(\mathbf{E}^2 - \mathbf{E} - 1)((a_n)) = (0).$$

In a general form, we have a polynomial $P(x)$ for an equation of the form $P(\mathbf{E})((a_n)) = (0)$. We can then find a general solution by using the facts that

$$(\mathbf{E} - \alpha)^k((n^j \alpha^n)) = (0)$$

for $0 \le j < k$, and

$$\{(n^j \alpha^n) \mid 0 \le j < k\}$$

is a linearly independent set, etc.

6. Of course, all of these are the easy cases. The hard ones are the nonlinear equations ...

7. At times, this reminds me of a comment made by my Ph.D. thesis advisor, after I had finished the proof of my main result for all odd primes. He said there were three ways to think about it:

   (a) I had handled infinitely many cases, and omitted only one, the even prime 2.

   (b) I had done half the cases. I had handled all the odd primes, but none of the even ones.

   (c) I had dealt with all the uninteresting cases, but not the single interesting case :-)

Exercises: Linear operators ←

1. Find a linear operator S such that

$$S^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

2. If S is the operator

$$S = \begin{bmatrix} 0 & -\frac{\pi}{4} \\ \frac{\pi}{4} & 0 \end{bmatrix},$$

   what is $\exp(S)$?

3. Solve the system of equations:

$$\begin{array}{rcrcrcl} 3x_1 & + & x_2 & - & 2x_3 & = & 3 \\ x_1 & - & 2x_2 & - & x_3 & = & 1 \\ 2x_1 & - & 4x_2 & - & x_3 & = & 3 \end{array}$$

4. Solve the differential equation:

$$\frac{d^5 f}{dx^5} - 7\frac{d^4 f}{dx^4} + 23\frac{d^3 f}{dx^3} - 45\frac{d^2 f}{dx^2} + 48\frac{df}{dx} - 20f$$

$$= 26\cos(x) - 18\sin(x)$$

   Hint:

$$x^5 - 7x^4 + 23x^3 - 45x^2 + 48x - 20$$

$$= (x-1)(x-2)^2(x^2 - 2x + 5)$$

5. Find the general solution to the difference equation

$$(\mathbf{E}^2 - \mathbf{E} - 1)((a_n)) = (0)$$

6. Find the general solution to the difference equation

$$(\mathbf{E}^4 + 2\mathbf{E}^2 + 1)((a_n)) = (0)$$

7. Verify that

   (a) D, the derivative, is a linear operator.

   (b) $(\mathrm{D} - r)^k(x^{k-1}e^{rx}) = 0$

$$\begin{array}{rcl} (\mathrm{D}^2 - 2s\mathrm{D} + s^2 + t^2)^k(x^{k-1}e^{sx}\cos(tx)) & = & 0 \\ (\mathrm{D}^2 - 2s\mathrm{D} + s^2 + t^2)^k(x^{k-1}e^{sx}\sin(tx)) & = & 0 \end{array}$$

   (c) $\mathbf{E}$, the discrete increment, is a linear operator.

   (d) $(\mathbf{E} - \alpha)^k((n^j\alpha^n)) = (0)$ for $0 \le j < k$.

8. What happens in nonlinear cases? Sometimes they are manageable, sometimes not.

   (a) Solve the differential equation

$$\mathrm{D}f = b * f * (1 - f).$$

(b) What can be said about the difference equation

$$\mathbf{E}((a_n)) = (b * a_n * (1 - a_n))?$$

Note: this is often called the logistics equation.

Normed linear spaces ←

- A **norm** on a real or complex vector space V is a function

$$|| \cdot || : V \to \mathbb{R}$$

with the properties, for $v, v_1, v_2 \in V$,

1. $||v|| \geq 0$
2. $||v|| = 0$ if and only if $u = \vec{0}$
3. $||av|| = |a| \ ||v||$
4. $||v_1 + v_2|| \leq ||v_1|| + ||v_2||$.

A space with such an associated function is called a **normed linear space**.

- Using the norm, we can define a **topology** on the space, and can then talk about continuity of functions or operators, limits of sequences, etc.

  I won't go into this much here beyond a few examples, but good places to look are books on Hilbert Spaces and/or functional analysis. There are a few books indicated in the references.

- Using the norm, we can define a **metric** on V by: $d(v_1, v_2) = ||v_1 - v_2||$. Metrics have the properties:

1. $d(v_1, v_2) \geq 0$
2. $d(v_1, v_2) = 0$ iff $v_1 = v_2$
3. $d(v_1, v_2) = d(v_2, v_1)$
4. $d(v_1, v_2) \leq d(v_1, v_3) + d(v_3, v_2)$ (the triangle inequality).

- Some examples:

1. On $\mathbb{R}^n$ or $\mathbb{C}^n$, we have the norm

$$||(a_1, a_2, \ldots, a_n)||_2 = (|a_1|^2 + \ldots + |a_n|^2)^{1/2}.$$

This is usually called the **Euclidean** norm.

We can also think of this in terms of the **inner product** given by

$$\langle (a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n) \rangle = a_1\bar{b}_1 + a_2\bar{b}_2 + \cdots + a_n\bar{b}_n.$$

We then have $||v||_2 = \langle v, v \rangle^{1/2}$.

2. We can generalize, for $p > 0$, to

$$||(a_1, a_2, \ldots, a_n)||_p = (|a_1|^p + \ldots + |a_n|^p)^{1/p},$$

and

$$||(a_1, a_2, \ldots, a_n)||_\infty = \max(|a_1|, \ldots, |a_n|).$$

These are called the p-norms and the ∞-norm (or max-norm).

A fun little exercise is to draw the circle of radius 1 in $\mathbb{R}^2$ for each of these norms:

$$circle_p(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_p = 1\}$$

for $p > 0$, or for $p = \infty$.

One of these constitutes a "proof" that a square is a circle :-)

3. We can generalize this to $\mathbb{R}^\infty$ or $\mathbb{C}^\infty$:

$$\begin{aligned} ||(a_1, a_2, \dots)||_p &= \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}, \\ ||(a_1, a_2, \dots)||_\infty &= \sup_n(|a_n|). \end{aligned}$$

For this to make sense, we will have to limit ourselves to the subspaces where the sum converges:

$$\begin{aligned} l^p &= \{(a_1, a_2, \dots) \mid \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p} < \infty\}, \\ l^\infty &= \{(a_1, a_2, \dots) \mid \sup_n(|a_n|) < \infty\}. \end{aligned}$$

4. On $C^0(\mathbb{R})$ or $C^0(\mathbb{C})$, we can define

$$||f||_p = \left(\int |f|^p\right)^{1/p}$$

$$||f||_\infty = \sup_x(|f(x)|).$$

Again, we limit ourselves to the subspaces where these are $< \infty$, and call the corresponding spaces L^p or L^∞.

5. We can think each of these Euclidean norms, $|| \cdot ||_2$, as coming from an inner product:

$$\begin{aligned} ||(a_1, a_2, \dots)||_2 &= \langle (a_1, \dots), (a_1, \dots)\rangle^{1/2} \\ &= \left(\sum_{n=1}^{\infty} a_n \bar{a}_n\right)^{1/2} \\ ||f||_2 &= \langle f, f\rangle^{1/2} \\ &= \left(\int f\bar{f}\right)^{1/2}. \end{aligned}$$

When a complex Euclidean normed linear space is **complete** (that is, every Cauchy sequence converges), it is called a **Hilbert space**.

Exercises: Normed linear spaces ←

1. Verify that each of the examples actually are norms.

2. In $\mathbb{R}^2$, draw the unit circles

   (a) $circle_1(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_1 = 1\}$

   (b) $circle_{3/2}(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_{3/2} = 1\}$

   (c) $circle_2(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_2 = 1\}$

   (d) $circle_3(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_3 = 1\}$

   (e) $circle_\infty(1) = \{(x,y) \in \mathbb{R}^2 \mid ||(x,y)||_\infty = 1\}$

3. Suppose that V is a real or complex vector space. An inner product on V is a conjugate-bilinear function on V:

$$< \cdot, \cdot >: V \times V \to \mathbb{F}$$

where, for all $v_1, v_2, v \in V$, and $\alpha \in \mathbb{F}$,

$$\begin{aligned} < v, v > &\geq 0 \\ < v, v > &= 0 \text{ iff } v = 0 \\ < v_1 + v_2, v > &= < v_1, v > + < v_2, v > \\ < v_1, v_2 > &= \overline{< v_2, v_1 >} \\ < \alpha v_1, v_2 > &= \alpha < v_1, v_2 > . \end{aligned}$$

Show that for an inner product,

$$\begin{aligned} < v, v_1 + v_2 > &= < v, v_1 > + < v, v_2 > \\ < v_1, \alpha v_2 > &= \bar{\alpha} < v_1, v_2 > . \end{aligned}$$

Show that for a finite dimensional real or complex vector space V with basis $(v_1, v_2, \dots, v_n)$, the function $f : V \times V \to \mathbb{F}$ given by

$$f\left(\sum_i a_i v_i, \sum_i b_i v_i\right) = \sum_i a_i \overline{b_i}$$

is an inner product.

4. Recall that a linear functional on a vector space V is a linear map $f : V \to \mathbb{F}$. For a finite dimensional real or complex inner product space V, define the **dual space** of V to be the space

$$V^* = \{f : V \to \mathbb{F} \mid f \text{ is a linear functional}\}$$

Show that V^* is a vector space over $\mathbb{F}$, and that V and V^* are isomorphic to each other.

Hint: Show that every $f \in V^*$ corresponds with a function of the form

$$< v_f, \cdot >: V \to \mathbb{F}$$

for some $v_f \in V$.

Eigenvectors and eigenvalues ←

- Suppose T is a linear operator on V. Then $\lambda \in \mathbb{F}$ is called an **eigenvalue** or **characteristic value** of T if there exists $v \neq 0$ in V with $T(v) = \lambda v$. In this case we call v an **eigenvector** or **characteristic vector** of T.

- An equivalent definition is that λ is an eigenvalue of T if and only if $(T - \lambda I)v = 0$ for some $v \neq 0$. This follows from the fact that $0 = (T - \lambda I)v = T(v) - \lambda v$ if and only if $T(v) = \lambda v$. Note that this also means that $\ker(T - \lambda I) \neq \{0\}$. Thus, λ is an eigenvalue if and only if $(T - \lambda I)$ is not a monomorphism. In the finite dimensional case, this is equivalent to $(T - \lambda I)$ not being invertible.

- Now suppose that $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T (i.e., $\lambda_i \neq \lambda_j$ for $i \neq j$), with associated eigenvectors $v_1, \ldots, v_k$. Then $\{v_1, \ldots, v_k\}$ is a linearly independent set.

- This means in particular that if $T \in L(V)$ with $\dim(V) = n$, and T has n distinct eigenvalues λ_i with associated eigenvectors v_i, then the set $S = (v_1, \ldots, v_n)$ is a basis for V. Furthermore, the matrix representation of T with respect to the basis S is

$$[T] = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

or $[T]_{ii} = \lambda_i$, $[T]_{ij} = 0$ for $i \neq j$. This is also sometimes written as $[T] = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Eigenvalues and eigenvectors can thus give us a very simple representation for T.

- Not all linear operators have eigenvalues (or eigenvectors). For example, the linear operator $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$[T_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

is a (counter-clockwise) rotation around the origin through the angle θ. If θ is not an integral multiple of π, then T_θ has no eigenvalues.

- More generally, consider a non-zero linear operator $T : \mathbb{R}^2 \to \mathbb{R}^2$. If $0 \neq u \in \mathbb{R}^2$, then there must be a linear combination of $\{u, T(u), T^2(u)\}$ with

$$0 = T^2(u) + aT(u) + b(u) = P_u(T)(u).$$

If $P_u(T)$ factors into linear factors $P_u(T) = (T - \lambda_1 I)(T - \lambda_2 I)$ with $\lambda_1, \lambda_2 \in \mathbb{R}$, then λ_1 and λ_2 are eigenvalues of T. On the other hand, if $a^2 - 4b < 0$, then T has no eigenvalues (it is a general rotation in $\mathbb{R}^2$).

- On the other hand, every non-zero linear operator $T : \mathbb{C}^n \to \mathbb{C}^n$ has an eigenvalue. We can see this from the fact that, for any $u \neq 0$ in $\mathbb{C}^n$, the set of vectors $\{u, T(u), T^2(u), \ldots, T^n(u)\}$ must be a linearly dependent set, and hence there are $a_0, a_1, \ldots, a_n$ not all zero with

$$0 = \sum_{i=0}^{n} a_i T^i(u) = P_u(T)(u).$$

  We know that we can factor the polynomial $P_u(T)$ over $\mathbb{C}$ into a product of linear factors

$$P_u(T) = \prod_j (T - z_j I)^{k_j}(T).$$

  This linear operator is not a monomorphism (since $u \in \ker(P_u(T))$). It is a product of commutative factors, and hence at least one of the factors $(T - z_j I)$ is not a monomorphism. This says then that at least one of the z_j is an eigenvalue of T. We can proceed to find n eigenvalues and eigenvectors.

- For for every non-zero linear operator $T : \mathbb{C}^n \to \mathbb{C}^n$, there is a polynomial $P_T(T) = \sum_i a_i T^i$ of degree n, called the **characteristic polynomial** of T, with:

  1. $P_T(T) = 0$. That is, $P_T(T)(u) = 0$ for all $u \in \mathbb{C}^n$.

  2. The linear factors of $P_T(T)$ are each of the form $(T - \lambda_j I)$ for some eigenvalue λ_j. Also, $a_n = 1$. That is,
$$P_T(T) = \prod_j (T - \lambda_j I)^{k_j}(T).$$

  3. For each eigenvalue λ_j, there is an eigenvector u_j. If $k_j > 1$, there is a linearly independent set of k_j eigenvectors for λ_j. The set of n eigenvectors $\{u_i\}$ is a linearly independent set, and thus is a basis for $\mathbb{C}^n$. As above, the matrix representation of T with respect to this basis is $[T] = \text{diag}(\lambda_1, \ldots, \lambda_n)$ (with repetitions as necessary).

- A non-zero linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ also has a characteristic polynomial of degree n. The big difference for the real case is that the polynomial will factor into a combination of linear and quadratic factors. There is then a basis consisting partially of eigenvectors and partially of pairs of basis vectors for two-dimensional subspaces on which T is a rotation. The matrix of T with respect to this basis then looks like $[T] = \text{diag}(\lambda_1, \ldots, \lambda_k, A_{k+1}, \ldots, A_{(n-k)/2})$ where each A_i is a two-by-two matrix without eigenvalues.

- The set of eigenvalues of an operator is sometimes called the **spectrum** of the operator. In cases like $\mathbb{C}^n$ where the eigenvectors form a basis for the space, using such as basis is sometimes known as the **spectral decomposition** of the operator.

- Just briefly, let's glance at differential operators. $\text{D} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ has uncountably many eigenvalues. Each real number $a \in \mathbb{R}$ is an eigenvalue since

$$\text{D}(e^{ax}) = ae^{ax}.$$

  The corresponding eigenvectors are of course $f_a(x) = e^{ax}$.

  The differential operator D^2 also has uncountably many eigenvalues and eigenvectors since for $a > 0$,

$$\begin{aligned} \text{D}^2(\cos(\sqrt{a}x)) &= -a\cos(\sqrt{a}x), \\ \text{D}^2(\sin(\sqrt{a}x)) &= -a\sin(\sqrt{a}x), \text{ and} \\ \text{D}^2(e^{\sqrt{a}x}) &= ae^{\sqrt{a}x}. \end{aligned}$$

- In this context, here's an example which puts many of these pieces together. In quantum mechanics, a typical expression of Schrödinger's equation looks like

$$\left[-\frac{\hbar^2}{2m_e}\bigtriangledown^2 + V(x,y,z) - i\hbar\frac{\partial}{\partial t}\right]\Psi = 0.$$

This example is for an electron (with mass m_e) in a potential field $V(x,y,z)$.

The general solution of this operator equation is

$$\Psi(x,y,z,t) = \sum_{n=0}^{\infty} c_n \Psi_n(x,y,z)\exp\left(\frac{-iE_n t}{\hbar}\right)$$

where $\Psi_n(x,y,z)$ is an eigenfunction solution of the associated time independent Schrödinger equation, with E_n the corresponding eigenvalue. The inner product, giving a time dependent probability, looks like

$$P(t) = \int \Psi\bar{\Psi}dv.$$

Exercises: Eigenvectors and eigenvalues ←

1. Show that if v is an eigenvector of T corresponding with the eigenvalue λ, and $\alpha \in \mathbb{F}$ ($\alpha \neq 0$), then αv is also an eigenvector of T corresponding with λ. In particular, eigenvectors are not unique.

2. Define $T \in L(\mathbb{F}^2)$ by $T((x,y)) = (y,x)$. Find all the eigenvalues and eigenvectors of T.

3. Define $T \in L(\mathbb{F}^3)$ by $T((x,y,z)) = (-y,0,2z)$. Find all the eigenvalues and eigenvectors of T.

4. Define $T \in L(\mathbb{F}^\infty)$ by $T((a_1,a_2,\ldots)) = (a_2,a_3,\ldots)$ (i.e., T is the left shift operator). Find all the eigenvalues and eigenvectors of T.

5. Suppose $T \in L(V)$ is invertible. Show that $\lambda \neq 0$ is an eigenvalue of T if and only if λ^{-1} is an eigenvalue of T^{-1}.

6. Suppose $S,T \in L(V)$. Show that ST and TS have the same eigenvalues.

7. Give an example of an operator whose matrix has all zeros on the diagonal with respect to some basis, but which is invertible. Give an example of an operator whose matrix has all diagonal elements non-zero with respect to some basis, but which is singular (i.e., has no inverse).

8. Show that if $S,T \in L(V)$, S is invertible, and $P(x) \in \mathbb{F}[x]$, then $P(S^{-1}TS) = S^{-1}P(T)S$.

9. Suppose that V is a finite dimensional normed complex vector space, and T is an isometry of V (i.e., $||T(v)|| = ||v||$ for all $v \in V$). Show that every eigenvalue λ of T has $|\lambda| = 1$. Hint: show that there is a basis for V consisting of eigenvectors e_i, with $||e_i|| = 1$ for $1 \leq i \leq n$.

10. Let V be a complex vector space, and $T \in L(V)$. Let $P(x) \in \mathbb{C}[x]$. Show that $\alpha \in \mathbb{C}$ is an eigenvalue of $P(T)$ if and only if $\alpha = P(\lambda)$ for some eigenvalue λ of T.

11. Is the preceding true for real vector spaces? Why not? (Note: this is another example of why $\mathbb{C}$ is so much nicer a place to work than $\mathbb{R}$ ...)

    Is the preceding true if we replace $\mathbb{C}[x]$ with $\mathbb{C}^\infty[x]$, power series? If so, show it. If not, give a counterexample.

Change of basis ←

- Recall that the matrix associated with a linear transformation depends on the particular bases we use. In the case of a linear operator $T : V \to V$, it is typical to use the same base for V as domain and as codomain. However, as we have seen, a basis consisting of eigenvectors is particularly convenient, and hence it is useful to know how to change bases.

  If V is finite dimensional, with two bases $S_1 = (u_1, \ldots, u_n)$ and $S_2 = (v_1, \ldots, v_n)$, we can consider the matrices of T with respect to the mixed bases. We can indicate this by the various symbols $[T]^{S_1}, [T]^{S_2}, [T]^{S_1 S_2}$, and $[T]^{S_2 S_1}$, where, for example

$$\begin{aligned} T(u_i) &= \sum_j [T]^{S_1}_{ji} u_j, \text{ and} \\ T(u_i) &= \sum_j [T]^{S_1 S_2}_{ji} v_j. \end{aligned}$$

- We can look in particular at the matrix representations for the identity operator I. For ease of notation, let $[B] = [I]^{S_1 S_2}$, the matrix representation for I using S_1 for the domain and S_2 for the codomain. In particular, we then have

$$u_i = \sum_j [B]_{ji} v_j = B^j_i v_j.$$

  We also have that the matrix inverse $[B]^{-1} = [I]^{S_2 S_1}$ which takes us in the opposite direction. Putting the pieces together, we then have

$$[B][T]^{S_1} = [T]^{S_2}[B].$$

  That is, doing the operator T in the basis S_1 and then converting to the basis S_2 is the same as converting bases first, and then doing T in the second base.

  Another way to say this is:

$$[T]^{S_1} = [B]^{-1}[T]^{S_2}[B]$$

  or

$$[T]^{S_2} = [B][T]^{S_1}[B]^{-1}.$$

Exercises: Change of basis ←

1. Show that if A and B are operators on a finite dimensional vector space with $AB = I$, the identity operator, then $BA = I$, and so $B = A^{-1}$.

2. Suppose that T is an operator that has the same matrix with respect to every basis. Show that T must be some multiple of the identity operator I.

Trace and determinant ←

- Let V be a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, and $T : V \to V$ an operator on V. Let $P_T(T) = \sum_{i=0}^n a_i T^i$ be the characteristic polynomial of T. We can then define the **trace** of T by trace$(T) = a_{n-1}$, the coefficient of T^{n-1}.

  We can also define the trace of an $n \times n$ real or complex matrix A by trace$_m(A) = \sum_{i=1}^n A_{ii}$, the sum of the diagonal elements of A.

  These definitions are consistent, in the sense that trace$(T) =$ trace$_m([T])$, where $[T]$ is the matrix of T with respect to some basis for V. An exercise will be to show that it doesn't matter what basis we use (they all come out the same). Since these definitions are consistent, we will ordinarily write them both the same way, as trace().

- For $S, T : V \to V$, and $a, b \in \mathbb{F}$, the trace has the nice properties:

  1. trace$(aS+bT) = a$ trace$(T)+b$ trace(T). This says that the trace is a linear functional on $L(V)$.

  2. In the complex case, we have
  $$\text{trace}(T) = \sum_i \lambda_i$$
  where the λ_i are the eigenvalues of T.

  3. trace$(ST) =$ trace(TS).

- From these we can derive nice facts like:

  If we define the **commutator** of S and T by $[S, T] = ST - TS$, then we always have $[S, T] \neq I$, the identity operator.

- We can define the **determinant** of an operator T by det$(T) = (-1)^n a_0$, where a_0 is the constant term in the characteristic polynomial.

- The determinant has the nice properties:

  1. In the complex case, we have
  $$\det(T) = \prod_i \lambda_i$$
  where the λ_i are the eigenvalues of T.

  2. det$(T) \neq 0$ iff ker$(T) = \{0\}$ iff T is invertible.

  3. det$(ST) =$ det(S) det$(T) =$ det(TS).

  4. $P_T(z) = \det(zI - T)$. That is, the characteristic polynomial is the determinant of the generalized operator $(zI - T)$, where we think of z as a complex variable.

5. We can define the determinant of a square matrix $M = [a_{ij}]$ by
$$\det{}_m(M) = \sum_{p \in \text{perm}(n)} \text{sign}(p) \prod_i a_{p(i)i}$$
where perm(n) is the set of all permutations of the numbers $(1, 2, \ldots, n)$, sign(p) is the sign of the permutation, and $p(i)$ is the ith number in the permutation p.

   We then have:
$$\det{}_m([T]) = \det(T).$$
   As for the trace, since the two definitions are consistent, we will typically denote both determinants by the same symbol det().

   Showing that these two definitions are consistent is a fair amount of work (just looking at the formula for det$_m$() should give you some idea). You can look it up if you are interested.

6. There is another definition of determinant. Suppose V is a real or complex n-dimensional vector space. Then there is an alternating multi-linear functional
$$\det{}_v : V^n \to \mathbb{F}$$
   with the properties, for all $v_1, v_2, \ldots, v_n, v_{i_1}, v_{i_2} \in V$ and $a, b \in \mathbb{F}$:

   (a) det$_v(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) =$ $-$det$_v(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)$ (alternating)

   (b) det$_v(v_1, \ldots, av_{i_1} + bv_{i_2}, \ldots, v_n) =$ $a$det$_v(v_1, \ldots, v_{i_1}, \ldots, v_n) +$ $b$det$_v(v_1, \ldots, v_{i_2}, \ldots, v_n)$ (multi-linear)

   (c) det$_v(e_1, \ldots, e_n) = 1$ for a particular basis $\{e_1, \ldots, e_n\}$ for V. (uniqueness)

   We then have det$_m(A) =$ det$_v(A_1, \ldots, A_n)$, where A_k

is the kth column of the matrix A, considered as a vector.

We have that $\det(T) = \det_m([T]) = \det_v([T]_1, \ldots, [T]_n)$. The fact that there are three different versions of the same thing suggests that many people have worked on this topic, and that this topic occurs in a variety of contexts ...

7. Recall the discussion of norms on vector spaces. The Euclidean norms ($||v||_2$), which can be thought of as coming from an inner product ($< v_1, v_2 >$), have a nice relationship with the determinant:

$$||T(v)||_2 = ||v||_2 \text{ for all } v \in U$$

$$\text{iff } |\det(T)| = 1.$$

8. An operator which preserves the Euclidean norm ($||T(v)||_2 = ||v||_2$) is

called an **isometry**. When the scalar field is $\mathbb{R}$, these are called **orthogonal** operators. In the case of $\mathbb{C}$, they are called **unitary** operators. In the case of the quaternions, they are called **symplectic** operators.

9. The isometries on a normed linear space have the properties:

(a) If T is an isometry, then $\det(T) \neq 0$, and hence T has an inverse, T^{-1}, which is also an isometry. Of course, I, the identity operator, is an isometry.

(b) If T_1 and T_2 are isometries, then T_1T_2 is also an isometry.

(c) This means that the set of isometries forms a **group**. In the three cases described above, these are called the **orthogonal group**, the **unitary group**, and the **symplectic group**.

10. Another way to characterize the isometries (in the finite dimensional case, but also with appropriate generalizations in infinite dimensional cases) is that an operator is an isometry if 1.) its inverse exists, and 2.) the matrix of its inverse is given by the conjugate transpose. That is, if

$$[T^{-1}]_{ij} = \overline{[T]_{ji}}.$$

11. The determinant also plays an important role in change of variables formulas (in $\mathbb{R}^n$, for example).

Suppose $\Omega \subset \mathbb{R}^n$ and $\sigma : \Omega \to \mathbb{R}^n$. We will think of σ as a (local) coordinate system on Ω, or as a change of variables.

The **derivative** of σ at $\mathbf{x}$ is the unique operator T (if it exists) satisfying:

$$\lim_{y \to 0} \frac{||\sigma(x+y) - \sigma(x) - Ty||}{||y||} = 0.$$

If this operator T exists, σ is called **differentiable**, and T is denoted by $\sigma'(x)$. Note that $\sigma'(x) \in L(\mathbb{R}^n)$. We can write $\sigma(x) = (\sigma_1(x), \ldots, \sigma_n(x))$, and we can denote the partial derivative of σ_j with respect to the kth coordinate by $\mathrm{D}_k\sigma_j(x)$.

If σ is differentiable at x, then the matrix of $\sigma'(x)$ is given by:

$$[\sigma'(x)] = \begin{bmatrix} \mathrm{D}_1\sigma_1(x) & \ldots & \mathrm{D}_n\sigma_1(x) \\ \vdots & & \vdots \\ \mathrm{D}_1\sigma_n(x) & \ldots & \mathrm{D}_n\sigma_n(x) \end{bmatrix}$$

If we assume that Ω is a reasonable set (e.g., open, or measurable) and $f : \Omega \to \mathbb{R}$ is also reasonable (e.g., continuous, or measurable), then we have the change of variables formula:

$$\int_{\sigma(\Omega)} f(y)dy = \int_{\Omega} f(\sigma(x))|\det(\sigma'(x))|dx.$$

Exercises: Trace and determinant ←

1. Given two real or complex nxn matrices A and B, show that $\text{trace}_m(AB) = \text{trace}_m(BA)$.

2. Given an operator T on a real or complex finite dimensional vector space V, show that $\text{trace}_m([T])$ and $\det_m([T])$ are independent of the basis used for the matrix representation $[T]$. Hint: use the change of basis formula $[T]^{S_1} = [B]^{-1}[T]^{S_2}[B]$, the previous problem, and $\det_m(AB) = \det_m(BA)$.

3. Show that trace is linear, that is, $\text{trace}(aS + bT) = a\text{trace}(S) + b\text{trace}(T)$.

4. Show that $\text{trace}(T) = \sum_{i=1}^{n}[T]_{ii}$.

5. For two operators S and T, show that $[S,T] \neq I$. Recall that $[S,T] = ST - TS$.

6. Show by example that in general, $\text{trace}(ST) \neq \text{trace}(S)\text{trace}(T)$. In particular, find an operator T on a real vector space V with $\text{trace}(T^2) < 0$.

7. Suppose A is an nxn real matrix, $S \in L(\mathbb{R}^n)$ has matrix representation A with respect to some basis, and $T \in L(\mathbb{C}^n)$ has matrix representation A with respect to some basis. Show that $\text{trace}(S) = \text{trace}(T)$ and $\det(S) = \det(T)$.

8. Show that if T is an isometry on a finite dimensional normed vector space, then $|\det(T)| = 1$.

9. Show that if T is an operator on a complex vector space of dimension n, and $\alpha \in \mathbb{C}$, then $\det(\alpha T) = \alpha^n \det(T)$.

To top ←

References

[1] Axler, Sheldon, *Linear Algebra Done Right,* second edition, Springer-verlag, New York, 1997.

[2] Finney, Ross L., and Ostberg, Donald R., *Elementary Differential Equations with Linear Algebra*, Addison-Wesley, 1976.

[3] Hoffman, Kenneth, and Kunze, Ray, *Linear Algebra,* second edition, Engineering/Science/Mathematics, 1971.

[4] Naimark, M. A., *Normed Algebras,* Wolters-Noordhoff, Groningen, the Netherlands, 1974.

[5] Noble, Ben, and Daniel, James W., *Applied Linear Algebra,* third edition, Engineering/Science/Mathematics, 1988.

[6] Prugovečki, Eduard, *Quantum Mechanics in Hilbert Space,* second edition, Academic Press, New York, 1981.

[7] Rudin, Walter, *Functional Analysis,* McGraw-Hill, New York, 1973.

[8] Strang, Gilbert, *Linear Algebra and its Applications,* Academic Press, New York, 1976.

To top ←