

2.4 The Algebra of Linear Maps

So far we have studied one linear map at a time: its kernel, its image, and the count that ties them together. But linear maps rarely come alone. The map $p \mapsto p + p'$ is the identity plus differentiation, and rotating then projecting is one map followed by another. This section shows that linear maps can be added, scaled, and composed, that these operations obey almost all the rules of arithmetic, and that an operator can be plugged into a polynomial just as a square matrix can.

2.4.1 Adding and scaling linear maps

Look at three maps on $\mathbb{R}[x]_{\leq 2}$ that have already appeared in this chapter or will soon:

$$p \mapsto p + p', \quad p \mapsto 3p', \quad p \mapsto p' - 2p.$$

Each is built from two simpler maps, the identity id and differentiation D , by the recipe “apply both to p , then combine the outputs with the operations of the target space”. The first is “ $\text{id} + D$ ”, the second “ $3D$ ”, the third “ $D - 2\text{id}$ ”. Instead of checking linearity for each such combination separately, we would like to check it once for the recipe. The recipe itself deserves a name.

Linear maps into the same space can be added and scaled output by output, and the results are again linear maps.

Definition 2.4.1. The space of linear maps

Let V and W be vector spaces over **the same field** F . We write $\mathcal{L}(V, W)$ for the set of **all** linear maps $V \rightarrow W$, and $\mathcal{L}(V) := \mathcal{L}(V, V)$ for the set of linear operators on V (Definition 2.1.1). For $S, T \in \mathcal{L}(V, W)$ and $c \in F$, the **sum** $S + T$ and the **scalar multiple** cT are the functions $V \rightarrow W$ defined by

$$(S + T)(v) := Sv + Tv, \quad (cT)(v) := c(Tv) \quad \text{for every } v \in V.$$

As in Definition 2.1.1, we write Tv for $T(v)$ when no brackets are needed. In words: to evaluate $S + T$ at v , evaluate both maps at v and add the results **in** W ; to evaluate cT , evaluate T and scale the result in W . The operations of V play no role in the definition. That is why S and T must have the **same codomain**: the sum $Sv + Tv$ has to be computed in one space. Both maps must also be defined on the same V .

We have used these operations before without naming them: the operations on F^X in Example (The Standard Vector Spaces) (d) are defined by the same pointwise recipe, with values added in F instead of W .

It turns out that the recipe never leaves the world of linear maps, and that $\mathcal{L}(V, W)$ is itself a vector space. This is the first sign that maps between vector spaces are as good as vectors.

Theorem 2.4.1. $\mathcal{L}(V, W)$ Is a Vector Space

Let V and W be vector spaces over F . If $S, T \in \mathcal{L}(V, W)$ and $c \in F$, then $S + T$ and cT are linear. With the operations of Definition 2.4.1, $\mathcal{L}(V, W)$ is a vector space over F . Its zero vector is the zero map $v \mapsto 0_W$, and the negative of T is $(-1)T: v \mapsto -Tv$.

Idea. There are two layers. The set W^V of **all** functions $V \rightarrow W$, with the pointwise operations, is a vector space, for the same reason F^X is: each axiom is checked one input at a time, where it becomes an axiom of W . Then $\mathcal{L}(V, W)$ sits inside W^V , so the Subspace Test finishes the job, and its three conditions are exactly “the zero map is linear”, “a sum of linear maps is linear”, and “a multiple of a linear map is linear”.

Proof. Let W^V be the set of all functions $V \rightarrow W$, with the operations of Definition 2.4.1 applied to arbitrary functions. The verification in Example (The Standard Vector Spaces) (d) goes through word for word with W in place of F : two functions $V \rightarrow W$ are equal when they agree at every $v \in V$ (Definition 0.4.1), and at each v every axiom for W^V becomes the same axiom in W , with zero function $v \mapsto 0_W$ and negative $v \mapsto -f(v)$. Hence W^V is a vector space over F , and $\mathcal{L}(V, W) \subseteq W^V$. We apply Theorem 1.2.1.

- (1) The zero map is linear: $0_W = 0_W + 0_W$ and $0_W = c0_W$ by Theorem 1.1.6. So $0 \in \mathcal{L}(V, W)$.
- (2) Let $S, T \in \mathcal{L}(V, W)$, $u, v \in V$, and $a \in F$. By the definition of $S + T$, the linearity of S and T , and the axioms of W ,

$$(S + T)(u + v) = Su + Sv + Tu + Tv = (S + T)u + (S + T)v,$$

$$(S + T)(av) = aSv + aTv = a(Sv + Tv) = a(S + T)v.$$

So $S + T \in \mathcal{L}(V, W)$.

- (3) Let $c \in F$ and $T \in \mathcal{L}(V, W)$. Then $(cT)(u + v) = c(Tu + Tv) = (cT)u + (cT)v$, and $(cT)(av) = c(aTv) = (ca)Tv = (ac)Tv = a(cT)v$, using the axioms of W and commutativity of multiplication in F . So $cT \in \mathcal{L}(V, W)$.

By Theorem 1.2.1, $\mathcal{L}(V, W)$ is a subspace of W^V , hence a vector space over F with the same zero and negatives. The negative of T in W^V is $v \mapsto -Tv = (-1)Tv$, by Theorem 1.1.7. This proves the theorem. ■

The step that used commutativity of F is worth a second look: cT is linear because c can be moved past the scalar a . This is one place where “ F is a field” is used silently.

Example. Sums of Matrix Maps and Other First Examples

For $\mathbf{A} \in M_{m \times n}(F)$, write $T_{\mathbf{A}}: F^n \rightarrow F^m$, $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$, for the matrix map of Example (Multiplication by a Matrix).

- (a) Show that $T_{\mathbf{A}} + T_{\mathbf{B}} = T_{\mathbf{A} + \mathbf{B}}$ and $cT_{\mathbf{A}} = T_{c\mathbf{A}}$ for $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$ and $c \in F$.
- (b) Describe $\mathcal{L}(V, W)$ when $V = \{0\}$, and when $W = \{0\}$.
- (c) On $\mathbb{R}[x]_{\leq 2}$, write $T(p) = p' - 2p$ in terms of id and D , and compute $T(1 + x^2)$.

Solution.

- (a) Let $\mathbf{x} \in F^n$. By [Definition 2.4.1](#) and distributivity ([Theorem 0.9.4](#)),

$$\begin{aligned}(T_{\mathbf{A}} + T_{\mathbf{B}})(\mathbf{x}) &= \mathbf{Ax} + \mathbf{Bx} = (\mathbf{A} + \mathbf{B})\mathbf{x} = T_{\mathbf{A}+\mathbf{B}}(\mathbf{x}), \\ (cT_{\mathbf{A}})(\mathbf{x}) &= c(\mathbf{Ax}) = (c\mathbf{A})\mathbf{x} = T_{c\mathbf{A}}(\mathbf{x}).\end{aligned}$$

The maps agree at every $\mathbf{x}$, so they are equal. So adding and scaling matrix maps is adding and scaling matrices.

- (b) If $V = \{0\}$, a linear map must send 0 to 0_W , so the zero map is the **only** element: $\mathcal{L}(\{0\}, W) = \{0\}$. If $W = \{0\}$, every function $V \rightarrow \{0\}$ is the zero map, and it is linear, so again $\mathcal{L}(V, \{0\}) = \{0\}$. These degenerate cases matter because formulas like “ $\dim \mathcal{L}(V, W) = \dim V \cdot \dim W$ ”, which we prove in Chapter 4, must give 0 here, and they do.
- (c) $T = D - 2\text{id}$, which is linear by [Theorem 2.4.1](#). With $p = 1 + x^2$, $p' = 2x$, so $T(p) = 2x - 2 - 2x^2$.

Here is a nonexample by minimal change. Keep D on $\mathbb{R}[x]_{\leq 2}$, but replace id by the evaluation map $E(p) = p(0) \in \mathbb{R}$. Both maps are linear and both are defined on $\mathbb{R}[x]_{\leq 2}$. But D lands in $\mathbb{R}[x]_{\leq 2}$ and E lands in $\mathbb{R}$, so $D + E$ would require adding a polynomial to a number. The clause that fails is “ $S, T \in \mathcal{L}(V, W)$ ” with the **same** W : the sum is not defined.

2.4.2 Composition

Addition makes $\mathcal{L}(V, W)$ a vector space, but the operation that really matters for maps is doing one after another. For functions, Chapter 0 wrote this as $S \circ T$ ([Definition 0.4.4](#)). For linear maps we drop the circle.

Remark. Notation. For $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, we write $ST := S \circ T: U \rightarrow W$, so $(ST)u = S(Tu)$. **The map written on the right acts first.** For an operator $T \in \mathcal{L}(V)$, $T^2 = TT$ means “apply T twice”, never “square the output”.

Is ST linear, and how does composition interact with the sum? It turns out that composition behaves like a multiplication: associative, with identities, and distributive over addition on both sides.

Theorem 2.4.2. Composition of Linear Maps

Let U, V, W, X be vector spaces over F .

- (a) **(Linearity)** If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $ST \in \mathcal{L}(U, W)$.
- (b) **(Associativity)** If $T \in \mathcal{L}(U, V)$, $S \in \mathcal{L}(V, W)$, and $R \in \mathcal{L}(W, X)$, then $R(ST) = (RS)T$.
- (c) **(Identities)** If $T \in \mathcal{L}(U, V)$, then $T \text{id}_U = T = \text{id}_V T$.
- (d) **(Distributivity)** If $T, T_1, T_2 \in \mathcal{L}(U, V)$, $S, S_1, S_2 \in \mathcal{L}(V, W)$, and $c \in F$, then

$$\begin{aligned}(S_1 + S_2)T &= S_1T + S_2T, \\ S(T_1 + T_2) &= ST_1 + ST_2, \\ (cS)T &= c(ST) = S(cT).\end{aligned}$$

Idea. Parts (b) and (c) hold for all functions, so there is nothing new to do. Part (a) is a two-line check. The interesting point is in (d): the three identities are not equally cheap. In $(S_1 + S_2)T$ the map T acts first and the sum is taken afterwards, so the identity is just the definition of the sum. In $S(T_1 + T_2)$ the sum is taken **before** S acts, and pulling it out of S needs S to be linear. So we should watch where the linearity of the **left** factor is spent.

Proof.

(a) Let $u, u' \in U$ and $a \in F$. Since T and then S are linear,

$$ST(u + u') = S(Tu + Tu') = STu + STu', \quad ST(au) = S(aTu) = aSTu.$$

(b) and (c) hold for arbitrary functions, by Theorem 0.4.2 and the observation after Definition 0.4.5.

(c) Each side is a function $U \rightarrow W$, so we compare values at $u \in U$. By Definition 2.4.1,

$$(S_1 + S_2)Tu = S_1(Tu) + S_2(Tu) = (S_1T + S_2T)u,$$

with no linearity used. Next, since S is linear,

$$S(T_1 + T_2)u = S(T_1u + T_2u) = ST_1u + ST_2u = (ST_1 + ST_2)u.$$

Finally $(cS)(Tu) = c(STu)$ by definition, and $S(cTu) = c(STu)$ because S is linear. This proves the theorem. ■

In short, $\mathcal{L}(V)$ carries a vector space structure and a multiplication (composition) that is associative, has the unit id_V , distributes over addition, and lets scalars move freely through products. A vector space with such a multiplication is called an **algebra** over F ; $M_n(F)$ is the other algebra we know (Theorem 0.9.4). The two will turn out to be the same algebra in disguise when we write operators as matrices in Chapter 4.

Example. Rotate, Then Project

Let $R: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the rotation by 90° counterclockwise, $R(x, y) = (-y, x)$, and let $P: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $P(x, y) = (x, 0)$, be the projection onto the x -axis. Compute PR and RP .

Solution. In PR the rotation acts first:

$$PR(x, y) = P(-y, x) = (-y, 0), \quad RP(x, y) = R(x, 0) = (0, x).$$

At $(1, 0)$: $PR(1, 0) = (0, 0)$, while $RP(1, 0) = (0, 1)$. So $PR \neq RP$. Geometrically, PR rotates $\mathbf{e}_1$ onto the y -axis and then squashes it to $\mathbf{0}$; RP leaves $\mathbf{e}_1$ alone under the projection and then rotates it to $\mathbf{e}_2$. In matrix form, $P = T_{\mathbf{A}}$ and $R = T_{\mathbf{B}}$ with $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and

$$\mathbf{B} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \text{ and indeed } \mathbf{AB} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \mathbf{BA}.$$

Warning. Composition is not commutative, and the order is read right to left. In [Example \(Rotate, Then Project\)](#), $PR \neq RP$. It can be worse: for $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ and $S \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$, both $ST \in \mathcal{L}(\mathbb{R}^2)$ and $TS \in \mathcal{L}(\mathbb{R}^3)$ are defined, but they live on different spaces, so they cannot even be compared. And for $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$, $S \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^4)$, ST is defined but TS is not. The thing you do first goes on the **right**.

Quick check. Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and $S: \mathbb{R}^2 \rightarrow \mathbb{R}^4$ be linear. Which of ST and TS is defined, and between which spaces does it go?

Solution. $ST = S \circ T$ first applies $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$, then $S: \mathbb{R}^2 \rightarrow \mathbb{R}^4$, so $ST \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^4)$. TS would apply S first, landing in $\mathbb{R}^4$, where T is not defined; so TS is not defined.

Because of the missing commutativity, school identities about products must be rechecked. By distributivity ([Theorem 2.4.2 \(d\)](#)),

$$(S + T)^2 = (S + T)(S + T) = S^2 + ST + TS + T^2,$$

which equals $S^2 + 2ST + T^2$ only if $ST = TS$. With P and R as above, $PR + RP: (x, y) \mapsto (-y, x)$, while $2PR: (x, y) \mapsto (-2y, 0)$; so $(P + R)^2 \neq P^2 + 2PR + R^2$. This is the operator version of the matrix warning in Chapter 0, §9.

2.4.3 Operators, powers, and polynomials

Composition of operators on one space V never leaves $\mathcal{L}(V)$, so we may compose an operator with itself as often as we like. Repeated composition leads to powers, and combinations of powers to polynomials, exactly as for square matrices ([Definition 0.9.14](#)).

Definition 2.4.2. Powers and Polynomials of an Operator

Let V be a vector space over F and $T \in \mathcal{L}(V)$. The **powers** of T are defined recursively by

$$T^0 := \text{id}_V, \quad T^{k+1} := T^k T \quad (k \in \mathbb{N}).$$

For $p = a_0 + a_1x + \cdots + a_Nx^N \in F[x]$, the operator $p(T) \in \mathcal{L}(V)$ is

$$p(T) := a_0 \text{id}_V + a_1 T + a_2 T^2 + \cdots + a_N T^N.$$

In words: T^k applies T exactly k times, and T^0 does nothing. In $p(T)$, the **constant term becomes** $a_0 \text{id}_V$, because an operator cannot be added to a scalar. Each T^k is linear by [Theorem 2.4.2 \(a\)](#) and induction, and $p(T)$ is linear by [Theorem 2.4.1](#). As for matrices, $p(T)$ does not depend on how many zero coefficients we write, since $0 \cdot T^k$ is the zero map.

Plugging in an operator respects sums and products of polynomials, just as plugging in a matrix does.

Theorem 2.4.3. Polynomials in One Operator

Let $T \in \mathcal{L}(V)$ and $p, q \in F[x]$. Then:

- (a) $T^j T^k = T^{j+k}$ for all $j, k \in \mathbb{N}$;
- (b) $(p + q)(T) = p(T) + q(T)$ and $(pq)(T) = p(T)q(T)$;
- (c) $p(T)q(T) = q(T)p(T)$.

Idea. The proof of Theorem 0.9.9 used four matrix rules: associativity, identity, distributivity, and “scalars move through products”. Theorem 2.4.2 supplies exactly these four rules for operators. So the old proof transfers line by line. Part (c) is the surprise: T need not commute with other operators, but it commutes with every polynomial in itself.

Proof.

- (a) Fix j and use induction on k (Theorem 0.2.7). For $k = 0$, $T^j T^0 = T^j \text{id}_V = T^j$ by Theorem 2.4.2 (c). If $T^j T^k = T^{j+k}$, then by Definition 2.4.2 and associativity (Theorem 2.4.2 (b)),

$$T^j T^{k+1} = T^j (T^k T) = (T^j T^k) T = T^{j+k} T = T^{j+k+1}.$$

- (b) Let $p = \sum_{i=0}^m a_i x^i$ and $q = \sum_{j=0}^l b_j x^j$. The statement for sums holds because $(a_i + b_i)T^i = a_i T^i + b_i T^i$ in the vector space $\mathcal{L}(V)$ (Theorem 2.4.1). For products, the coefficient of x^k in pq is $\sum_{i+j=k} a_i b_j$ (Definition 0.8.2), so

$$\begin{aligned} (pq)(T) &= \sum_k \left(\sum_{i+j=k} a_i b_j \right) T^k = \sum_{i=0}^m \sum_{j=0}^l (a_i b_j) T^i T^j \\ &= \left(\sum_{i=0}^m a_i T^i \right) \left(\sum_{j=0}^l b_j T^j \right) = p(T) q(T), \end{aligned}$$

where the second equality uses (a), and the third uses both distributive laws and $(a_i T^i)(b_j T^j) = (a_i b_j) T^i T^j$, all from Theorem 2.4.2 (d).

- (c) Since $pq = qp$ in $F[x]$ (Theorem 0.8.1), part (b) gives $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$. This proves the theorem. ■

In practice, (b) says that **a factorization of p is a factorization of $p(T)$** : for instance $T^2 - 3T + 2\text{id}_V = (T - \text{id}_V)(T - 2\text{id}_V) = (T - 2\text{id}_V)(T - \text{id}_V)$.

Example. Differentiation Is Nilpotent

Let $n \geq 0$ and let $D \in \mathcal{L}(F[x]_{\leq n})$ be (formal) differentiation (Example (Differentiation)), $D(a_0 + a_1 x + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + na_n x^{n-1}$, where ka_k means a_k added to itself k times. Show that $D^{n+1} = 0$. Show also that $D^n \neq 0$ when $F = \mathbb{R}$.

Solution. For $0 \leq k \leq n$, $D(x^k) = kx^{k-1}$ if $k \geq 1$ and $D(1) = 0$. So D sends each x^k to a scalar multiple of x^{k-1} , or to 0 when $k = 0$. Applying this $k+1$ times, $D^{k+1}(x^k)$ is a scalar multiple of $D(1) = 0$, so $D^{k+1}(x^k) = 0$. Since $k \leq n$, also $D^{n+1}(x^k) = D^{n-k}(D^{k+1}(x^k)) = D^{n-k}(0) = 0$, by Theorem 2.4.3 (a).

Now let $p = a_0 + a_1x + \cdots + a_nx^n$. Since D^{n+1} is linear, Theorem 2.1.4 gives

$$D^{n+1}(p) = a_0D^{n+1}(1) + a_1D^{n+1}(x) + \cdots + a_nD^{n+1}(x^n) = 0.$$

Hence $D^{n+1} = 0$.

Over $\mathbb{R}$, $D^n(x^n) = n(n-1)\cdots 1 = n! \neq 0$, so $D^n \neq 0$. Over $\mathbb{F}_2$ with $n = 2$, however, $D(x^2) = 2x = 0$, and $D^2 = 0$ already; the field matters for the exact power, but never for $D^{n+1} = 0$.

An operator T with $T^k = 0$ for some $k \geq 1$ is called **nilpotent**. The example shows that a nonzero operator can have a power equal to zero, just as the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ can. Differentiation is the model nilpotent operator, and it returns when we study canonical forms.

Example. Operators That Satisfy a Polynomial

With P and R as in [Example \(Rotate, Then Project\)](#), find nonzero polynomials p and q with $p(P) = 0$ and $q(R) = 0$. Also, for an arbitrary $p \in F[x]$, compute $p(\text{id}_V)$ and $p(Z)$, where $Z \in \mathcal{L}(V)$ is the zero operator.

Solution. For the projection, $P^2(x, y) = P(x, 0) = (x, 0) = P(x, y)$, so $P^2 = P$ and $p = x^2 - x$ satisfies $p(P) = P^2 - P = 0$. By [Theorem 2.4.3](#) (b), this says $P(P - \text{id}) = 0$: whatever $P - \text{id}$ produces is killed by P , and indeed $(P - \text{id})(x, y) = (0, -y)$ lies on the y -axis.

For the rotation, $R^2(x, y) = R(-y, x) = (-x, -y)$, so $R^2 = -\text{id}$, and $q = x^2 + 1$ satisfies $q(R) = R^2 + \text{id} = 0$. Rotating twice by 90° is rotating by 180° . Since q has no real roots, q has no real linear factor, so the trick “factor q , then factor $q(R)$ ” gives nothing over $\mathbb{R}$.

For $p = a_0 + a_1x + \cdots + a_Nx^N$: every power of id_V is id_V , so $p(\text{id}_V) = (a_0 + a_1 + \cdots + a_N)\text{id}_V = p(1)\text{id}_V$. Every power Z^k with $k \geq 1$ is the zero map, but $Z^0 = \text{id}_V$, so $p(Z) = a_0\text{id}_V$, the scalar $p(0)$ times id_V . The second computation is the reason for the convention $T^0 = \text{id}_V$: with it, “the constant term” of $p(T)$ is what survives when T is the zero operator.

Warning. The constant term is $a_0\text{id}_V$, not a_0 . Writing “ $P^2 - P + 3$ ” for an operator is a type error: $P^2 - P$ is a map and 3 is a number. For $p = x^2 - x + 3$, $p(P) = P^2 - P + 3\text{id}$, which is 3id since $P^2 = P$. Forgetting the id leads to “ $p(P) = 3$ ”, which is not an operator at all.

Quick check. With P the projection above, compute $p(P)$ for $p = x^3 - 2x + 1$.

Solution. Since $P^2 = P$, also $P^3 = P^2P = PP = P$. So $p(P) = P^3 - 2P + \text{id} = P - 2P + \text{id} = \text{id} - P$, which is the map $(x, y) \mapsto (0, y)$: the projection onto the y -axis.

Polynomial relations such as $P^2 = P$, $R^2 = -\text{id}$, and $D^{n+1} = 0$ carry real information about an

operator. In the next section we use one to write down an inverse without solving any equations, and later in the book the polynomials that kill an operator lead to the minimal polynomial and to canonical forms.

2.4.4 Exercises

2.4.4.1 A. Check your understanding

Exercise. A1

- (a) For $S, T \in \mathcal{L}(V, W)$ and $c \in F$, define $S + T$ and cT . Why must S and T have the same codomain?
- (b) True or false: for all $S, T \in \mathcal{L}(V)$, $ST = TS$. Justify your answer.
- (c) Let $T \in \mathcal{L}(\mathbb{R}^5, \mathbb{R}^2)$ and $S \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^5)$. Which of ST , TS are defined, and in which spaces do they lie?
- (d) For $T \in \mathcal{L}(V)$ and $p = 4$ (a constant polynomial), what is $p(T)$?
- (e) True or false: for all $S, T \in \mathcal{L}(V)$, $(S + T)^2 = S^2 + 2ST + T^2$. Justify your answer.

Solution.

- (a) $(S + T)(v) = Sv + Tv$ and $(cT)(v) = cTv$ for every $v \in V$ ([Definition 2.4.1](#)). The sum $Sv + Tv$ must be computed in a single vector space, so both values must lie in the same W .
- (b) False. In [Example \(Rotate, Then Project\)](#), $PR(1, 0) = (0, 0)$ but $RP(1, 0) = (0, 1)$.
- (c) Both are defined. ST applies T first: $ST \in \mathcal{L}(\mathbb{R}^5)$. TS applies S first: $TS \in \mathcal{L}(\mathbb{R}^2)$.
- (d) $p(T) = 4 \text{id}_V$, by [Definition 2.4.2](#).
- (e) False. By [Theorem 2.4.2](#) (d), $(S + T)^2 = S^2 + ST + TS + T^2$, which differs from $S^2 + 2ST + T^2$ by $TS - ST$. With $S = P$, $T = R$ of [Example \(Rotate, Then Project\)](#), $RP - PR: (x, y) \mapsto (y, x)$ is not the zero map, so the identity fails.

2.4.4.2 B. Practice

Exercise. B1: Computing compositions

Let $S, T \in \mathcal{L}(\mathbb{R}^2)$ be given by $S(x, y) = (x + y, y)$ and $T(x, y) = (y, x)$.

- (a) Compute ST , TS , and $2S - T$.
- (b) Compute S^2 and S^k for every $k \in \mathbb{N}$.
- (c) Hence show that $(S - \text{id})^2 = 0$.

Solution.

- (a) Applying the right-hand map first,

$$ST(x, y) = S(y, x) = (x + y, x), \quad TS(x, y) = T(x + y, y) = (y, x + y).$$

So $ST \neq TS$; for instance $ST(1, 0) = (1, 1)$ and $TS(1, 0) = (0, 1)$. By [Definition 2.4.1](#),

$$(2S - T)(x, y) = (2x + 2y, 2y) - (y, x) = (2x + y, 2y - x).$$

- (b) $S^2(x, y) = S(x + y, y) = (x + 2y, y)$. We claim $S^k(x, y) = (x + ky, y)$ for all $k \in \mathbb{N}$, by induction on k . For $k = 0$, $S^0 = \text{id}$ sends (x, y) to $(x + 0y, y)$. If the claim holds for k , then

$$S^{k+1}(x, y) = S^k(S(x, y)) = S^k(x + y, y) = (x + y + ky, y) = (x + (k + 1)y, y),$$

where the first equality is [Definition 2.4.2](#), $S^{k+1} = S^k S$.

- (c) $(S - \text{id})(x, y) = (y, 0)$, so $(S - \text{id})^2(x, y) = (S - \text{id})(y, 0) = (0, 0)$. Hence $(S - \text{id})^2 = 0$, that is, $S - \text{id}$ is nilpotent. Expanding with [Theorem 2.4.3](#) (b), $S^2 - 2S + \text{id} = 0$, which matches (b): $(x + 2y, y) - 2(x + y, y) + (x, y) = (0, 0)$.

Exercise. B2: Operators that preserve a subspace

Let U be a subspace of V , and let

$$\mathcal{A}_U := \{T \in \mathcal{L}(V) : Tu \in U \text{ for every } u \in U\}.$$

Prove that $\mathcal{A}_U$ is a subspace of $\mathcal{L}(V)$, that $\text{id}_V \in \mathcal{A}_U$, and that $ST \in \mathcal{A}_U$ whenever $S, T \in \mathcal{A}_U$.

Solution. We use Theorem 1.2.1 in the vector space $\mathcal{L}(V)$ ([Theorem 2.4.1](#)).

- (1) The zero map sends every $u \in U$ to 0, and $0 \in U$ because U is a subspace. So $0 \in \mathcal{A}_U$.
- (2) Let $S, T \in \mathcal{A}_U$ and $u \in U$. Then $Su \in U$ and $Tu \in U$, so $(S + T)u = Su + Tu \in U$, since U is closed under addition. Therefore $S + T \in \mathcal{A}_U$.
- (3) Let $c \in F$ and $T \in \mathcal{A}_U$. For $u \in U$, $(cT)u = cTu \in U$, since U is closed under scalar multiplication. Therefore $cT \in \mathcal{A}_U$.

Hence $\mathcal{A}_U$ is a subspace of $\mathcal{L}(V)$. Next, $\text{id}_V u = u \in U$ for every $u \in U$, so $\text{id}_V \in \mathcal{A}_U$. Finally let $S, T \in \mathcal{A}_U$ and $u \in U$. Then $Tu \in U$, and applying the hypothesis on S to the vector $Tu \in U$ gives $STu = S(Tu) \in U$. This shows $ST \in \mathcal{A}_U$.

Exercise. B3: A polynomial that kills D

Let $D \in \mathcal{L}(\mathbb{R}[x]_{\leq 2})$ be differentiation.

- (a) Find a nonzero polynomial p with $p(D) = 0$.
- (b) Show that no nonzero polynomial q of degree at most 2 satisfies $q(D) = 0$.
- (c) Hence, or otherwise, compute $(\text{id} + D)^3$ as a combination of id, D, D^2 , and evaluate $(\text{id} + D)^3(x^2)$.

Solution.

- (a) By [Example \(Differentiation Is Nilpotent\)](#) with $n = 2$, $D^3 = 0$. So $p = x^3$ works.
- (b) Let $q = a + bx + cx^2$ with $q(D) = a \text{id} + bD + cD^2 = 0$. Applying this operator to x^2 ,

with $D(x^2) = 2x$ and $D^2(x^2) = 2$,

$$0 = ax^2 + 2bx + 2c.$$

A polynomial is zero exactly when all its coefficients are, so $a = 0$, $2b = 0$, and $2c = 0$. Hence $a = b = c = 0$ and $q = 0$.

- (c) The operator $\text{id} + D$ is $q(D)$ for $q = 1 + x$. Applying the product rule of [Theorem 2.4.3](#) (b) twice, $(\text{id} + D)^3 = q(D)q(D)q(D) = (q^3)(D)$, and $q^3 = 1 + 3x + 3x^2 + x^3$. So

$$(\text{id} + D)^3 = \text{id} + 3D + 3D^2 + D^3 = \text{id} + 3D + 3D^2,$$

using $D^3 = 0$ from (a). Therefore $(\text{id} + D)^3(x^2) = x^2 + 3 \cdot 2x + 3 \cdot 2 = x^2 + 6x + 6$.

2.4.4.3 C. Going deeper

Exercise. C1: The Heisenberg relation

Let F be a field, let $D \in \mathcal{L}(F[x])$ be formal differentiation, $D(\sum_k a_k x^k) = \sum_{k \geq 1} k a_k x^{k-1}$ (Example (Differentiation)), and let $S \in \mathcal{L}(F[x])$ be multiplication by x , $S(p) = xp$ (the map M of Example (Multiplication by x)).

- (a) Prove that $DS - SD = \text{id}$.
- (b) Let $T \in \mathcal{L}(F^n, F^m)$. Prove that $T = T_{\mathbf{A}}$, where $\mathbf{A} \in M_{m \times n}(F)$ is the matrix with columns $T\mathbf{e}_1, \dots, T\mathbf{e}_n$.
- (c) Now suppose that F has **characteristic** 0 (Definition 0.7.2), and let $n \geq 1$. Deduce that there are no operators $D', S' \in \mathcal{L}(F^n)$ with $D'S' - S'D' = \text{id}$.
- (d) Show that the assumption on the characteristic in (c) cannot be dropped.

Hint: for (c), use (b) and consider traces.

Solution.

- (a) Both sides are linear ([Theorem 2.4.1](#), [Theorem 2.4.2](#)). We first compare them on x^k , $k \in \mathbb{N}$. We have $DS(x^k) = D(x^{k+1}) = (k+1)x^k$. If $k \geq 1$, $SD(x^k) = x \cdot kx^{k-1} = kx^k$; if $k = 0$, $SD(1) = x \cdot 0 = 0 = 0 \cdot x^0$. In both cases

$$(DS - SD)(x^k) = (k+1)x^k - kx^k = x^k = \text{id}(x^k).$$

Every $p \in F[x]$ is a finite combination $p = \sum_k a_k x^k$, so by Theorem 2.1.4, $(DS - SD)(p) = \sum_k a_k (DS - SD)(x^k) = \sum_k a_k x^k = p$. Hence $DS - SD = \text{id}$.

- (b) Let $\mathbf{x} = (x_1, \dots, x_n) = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n \in F^n$. By Theorem 2.1.4 and Theorem 0.9.3,

$$T\mathbf{x} = x_1T\mathbf{e}_1 + \dots + x_nT\mathbf{e}_n = \mathbf{A}\mathbf{x} = T_{\mathbf{A}}(\mathbf{x}).$$

Since $\mathbf{x}$ was arbitrary, $T = T_{\mathbf{A}}$.

- (c) Suppose, for a contradiction, that $D', S' \in \mathcal{L}(F^n)$ satisfy $D'S' - S'D' = \text{id}$. By (b), $D' = T_{\mathbf{A}}$ and $S' = T_{\mathbf{B}}$ for some $\mathbf{A}, \mathbf{B} \in M_n(F)$, and $\text{id} = T_{\mathbf{I}_n}$. By associativity of matrix multiplication, $T_{\mathbf{A}}T_{\mathbf{B}}(\mathbf{x}) = \mathbf{A}(\mathbf{B}\mathbf{x}) = (\mathbf{A}\mathbf{B})\mathbf{x}$, so $T_{\mathbf{A}}T_{\mathbf{B}} = T_{\mathbf{A}\mathbf{B}}$, and similarly $T_{\mathbf{B}}T_{\mathbf{A}} = T_{\mathbf{B}\mathbf{A}}$. With [Example \(Sums of Matrix Maps and Other First Examples\)](#) (a),

$$T_{\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}} = T_{\mathbf{A}}T_{\mathbf{B}} - T_{\mathbf{B}}T_{\mathbf{A}} = \text{id} = T_{\mathbf{I}_n}.$$

Evaluating at $\mathbf{e}_j$ gives equal j -th columns (Theorem 0.9.3), so $\mathbf{AB} - \mathbf{BA} = \mathbf{I}_n$. Taking traces and using Theorem 0.9.6,

$$n \cdot 1 = \operatorname{tr} \mathbf{I}_n = \operatorname{tr}(\mathbf{AB}) - \operatorname{tr}(\mathbf{BA}) = 0.$$

Since $n \geq 1$ and F has characteristic 0, $n \cdot 1 \neq 0$ by Definition 0.7.2. This is a contradiction, so no such D', S' exist. In words: over $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or any field of characteristic 0, the relation $DS - SD = \operatorname{id}$ of (a) has no analogue on F^n , $n \geq 1$. Since every n -dimensional space is isomorphic to F^n , as we show in the next section, the same holds on every nonzero finite-dimensional space. (On the zero space $\operatorname{id} = 0$, so the relation holds trivially there; that is why $n \geq 1$.)

- (d) Over $\mathbb{F}_2$ (characteristic 2), take $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ as in Exercise (C1: $\mathbf{AB} - \mathbf{BA} = \mathbf{I}$ is impossible, or is it?). There $\mathbf{AB} - \mathbf{BA} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2$, because $-1 = 1$ in $\mathbb{F}_2$. By the computation in (c), $D' = T_{\mathbf{A}}$ and $S' = T_{\mathbf{B}}$ satisfy $D'S' - S'D' = T_{\mathbf{AB}-\mathbf{BA}} = \operatorname{id}$ on $\mathbb{F}_2^2$. So the conclusion of (c) fails without the characteristic assumption; the proof breaks exactly at $n \cdot 1 \neq 0$, since $2 \cdot 1 = 0$ in $\mathbb{F}_2$.

Exercise. C2: Commuting nilpotent operators

Let $S, T \in \mathcal{L}(V)$.

- Suppose $ST = TS$. Prove that $(S + T)^n = \sum_{k=0}^n \binom{n}{k} S^k T^{n-k}$ for every $n \in \mathbb{N}$, where $\binom{n}{k} S^k T^{n-k}$ means $S^k T^{n-k}$ added to itself $\binom{n}{k}$ times.
- Deduce that if $ST = TS$, $S^a = 0$, and $T^b = 0$ for some $a, b \geq 1$, then $(S + T)^{a+b-1} = 0$.
- Show that (b) fails without $ST = TS$: give nilpotent $S, T \in \mathcal{L}(\mathbb{R}^2)$ such that $S + T$ is not nilpotent.

Hint: for (c), try maps that send one standard basis vector to the other and the other to 0.

Solution.

- First, $T^j S = ST^j$ for all $j \in \mathbb{N}$, by induction on j : it holds for $j = 0$, and if $T^j S = ST^j$, then $T^{j+1} S = T^j(TS) = T^j(ST) = (T^j S)T = (ST^j)T = ST^{j+1}$, using associativity (Theorem 2.4.2 (b)) and Theorem 2.4.3 (a) in the form $T^{j+1} = T^j T$. Now induct on n . For $n = 0$ both sides are id_V . Suppose the formula holds for n . By Theorem 2.4.2 (d),

$$(S + T)^{n+1} = (S + T)^n (S + T) = \sum_{k=0}^n \binom{n}{k} S^k T^{n-k} S + \sum_{k=0}^n \binom{n}{k} S^k T^{n-k+1}.$$

By the first paragraph, $S^k T^{n-k} S = S^k (T^{n-k} S) = S^k S T^{n-k} = S^{k+1} T^{n-k}$. Shifting the index in the first sum, the coefficient of $S^k T^{n+1-k}$ becomes $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$ (Pascal's rule, with $\binom{n}{-1} = \binom{n}{n+1} = 0$). This is the formula for $n + 1$.

- By (a) with $n = a + b - 1$, $(S + T)^{a+b-1}$ is a sum of multiples of $S^k T^{a+b-1-k}$, $0 \leq k \leq a + b - 1$. If $k \geq a$, then $S^k = S^{k-a} S^a = 0$ by Theorem 2.4.3 (a), so the term is 0 by Theorem 2.4.2 (d). If $k \leq a - 1$, then $a + b - 1 - k \geq b$, so $T^{a+b-1-k} = 0$ likewise. Every term vanishes, so $(S + T)^{a+b-1} = 0$.

- (c) Let $S(x, y) = (y, 0)$ and $T(x, y) = (0, x)$. Then $S^2(x, y) = S(y, 0) = (0, 0)$ and $T^2(x, y) = T(0, x) = (0, 0)$, so both are nilpotent. But $(S + T)(x, y) = (y, x)$, so $(S + T)^2(x, y) = (x, y)$, that is, $(S + T)^2 = \text{id}$. Then $(S + T)^{2k} = \text{id} \neq 0$ and $(S + T)^{2k+1} = S + T \neq 0$ for every k , so $S + T$ is not nilpotent. Here $ST(x, y) = (x, 0) \neq (0, y) = TS(x, y)$, so the hypothesis of (b) indeed fails.