Chapter II. Linear Operators $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{W}$.

II.1 Generalities.

A map $T : V \rightarrow W$ between vector spaces is a **linear operator** if for any $v, v_1, v_2 \in V$ and $\lambda \in \mathbb{K}$

1. Scaling operations are preserved: $T(\lambda \cdot v) = \lambda \cdot T(v)$
2. Sums are preserved: $T(v_1 + v_2) = T(v_1) + T(v_2)$

This is equivalent to saying

$$T\Big(\sum_i \lambda_i v_i\Big) = \sum_i \lambda_i T(v_i) \qquad \text{in } W$$

for all finite linear combinations of vectors in V. A trivial example is the **zero operator** $T(v) = 0_W$, for every $v \in V$. If $W = V$ the **identity operator**, $\mathrm{id}_V : V \rightarrow V$ is given by $\mathrm{id}(v) = v$ for all vectors. Some basic properties of any linear operator $T : V \rightarrow W$ are:

1. $T(0_V) = 0_W$. [Proof: $T(0_V) = T(0 \cdot 0_V) = 0 \cdot T(0_V) = 0_W$.]
2. $T(-v) = -T(v)$. [Proof: $T(-v) = T((-1) \cdot v) = (-1) \cdot T(0_V) = 0_W$.]
3. A linear operator is determined by its action on any set S of vectors that span V. If $T_1, T_2 : V \rightarrow W$ are linear operators and

   $$T_1(s) = T_2(s) \quad \text{ for all } \quad s \in S,$$

   then $T_1 = T_2$ everywhere on V. [Proof: Any $v \in V$ is a finite linear combination $v = \sum_i c_i s_i$; then $T_1(v) = \sum_i c_i T_i(s_i) = T_2(v)$.]

1.1. Exercise. If $S \subseteq V$ and $T : V \rightarrow W$ is a linear operator prove that

$$T\big(\mathbb{K}\text{-span}\{S\}\big) \text{ is equal to } \mathbb{K}\text{-span}\{T(S)\}.$$

1.2. Definition. *We write* $\mathrm{Hom}_{\mathbb{K}}(V, W)$ *for the space of linear operators* $T : V \rightarrow W$. *This becomes a vector space over* $\mathbb{K}$ *if we define*

1. $(T_1 + T_2)(v) = T_1(v) + T_2(v)$;
2. $(\lambda \cdot T)(v) = \lambda \cdot (T(v))$.

for any $v \in V$, $\lambda \in \mathbb{K}$. *The vector space axioms are easily verified. The zero element in* $\mathrm{Hom}(V, W)$ *is the zero operator:* $0(v) = 0_W$ *for every* $v \in V$. *The additive inverse* $-T$ *is is the operator* $-T(v) = (-1) \cdot T(v) = T(-v)$, *which is also a scalar multiple* $-T = (-1)T$ *of* T. $\square$

If $V = W$ we can also define the **composition product** $S \circ T$ of two operators,

$$(S \circ T)(v) = S(T(v)) \qquad \text{for all } v \in V$$

This makes $\mathrm{Hom}_{\mathbb{K}}(V) = \mathrm{Hom}_{\mathbb{K}}(V, V)$ a noncommutative associative algebra with identity $I = \mathrm{id}_V$.

A linear operator $T : V \rightarrow W$ over $\mathbb{K}$ determines two important vector subspaces, the **kernel** $K(T) = \ker(T)$ in the initial space V and the **range** $R(T) = \mathrm{range}(T)$ in the target space W.

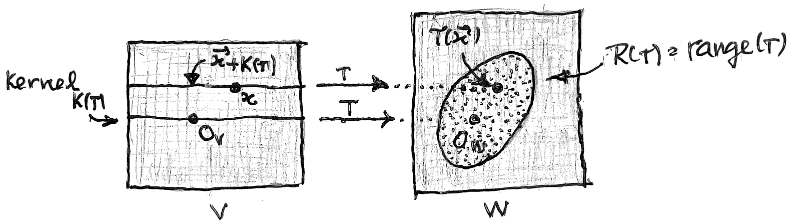

Figure 2.1. A linear map $T : V \to W$ sends all points in a coset $x + K(T)$ of the kernel to a single point $T(x)$ in W. Different cosets map to different points, and all images land within the range $R(T)$. The zero coset $0_V + K(T)$ collapses to the origin 0_W in W.

1. $K(T) = \ker(T) = \{v \in V : T(v) = 0_W\}$. The dimension of this space is often referred to as $\text{Nullity}(T)$.
2. $R(T) = \text{range}(T)$ is the image set $T(V) = \{T(v) : v \in V\}$. Its dimension is the **rank**

$$\text{rank}(T) = \dim_{\mathbb{K}}\big(\text{range}(T)\big) \ ,$$

which is often abbreviated as $\text{rk}(T)$.

1.3. Exercise. Show that $\ker(T)$ and $\text{range}(T)$ are vector subspaces of V and W respectively.

We often have to decide whether a linear map is one-to-one, onto, or a bijection. For surjectivity, we must compute the range $R(T)$; determining whether T is one-to-one is easier, and amounts to computing the kernel $K(T)$. The diagram Figure 2.1 illustrates the general behavior of any linear operator. Each coset $v + K(T)$ of the kernel gets mapped to a single point in W because

$$T(v + K(T)) = T(v) + T(K(T)) = T(v) + 0_W = T(v) \quad ,$$

and distinct cosets go to different points in W. All points in V map into the range $R(T) \subseteq W$.

1.4. Lemma. *A linear operator $T : V \to W$ is one-to-one if and only if* $\ker(T) = 0$.

Proof: $(\Leftarrow)$. If $T(v_1) = T(v_2)$ for $v_1 \neq v_2$, then $0 = T(v_1) - T(v_2) = T(v_1 - v_2)$, so $v_2 - v_1 \neq 0$ is in $\ker(T)$ and the kernel is nontrivial. We have just proved the "contrapositive" $\neg\big(T \text{ is one-to-one}\big) \Rightarrow \neg\big(K(T) = \{0\}\big)$ of the statment $(\Leftarrow)$ we want, but the two are logically equivalent.
Moral: If you want to prove $(P \Rightarrow Q)$ it is sometimes easier to prove the equivalent *contrapositive statement* $(\neg Q \Rightarrow \neg P)$, as was the case here.

Proof: $(\Rightarrow)$. Suppose T is one-to-one. If $v \neq 0_V$ then $Tv \neq T(0_V) = 0_W$ so $v \notin K(T)$ and $K(T)$ reduces to $\{0\}$. $\square$

The following important result is closely related to Theorem 5.7 (Chapter I) for quotient spaces.

1.5. Theorem (The Dimension Theorem). *If $T : V \to W$ is a linear operator and V is finite dimensional, the range $R(T)$ is finite dimensional and is related to the kernel $K(T)$ via*

$$\dim(R(T)) + \dim(K(T)) = \dim(V) \tag{7}$$

In words, "rank + nullity = dimension of the initial space V." This can also be expressed in short form by writing $|R(T)| + |K(T)| = |V|$.

Proof: The kernel is finite dimensional because $K(T) \subseteq V \Rightarrow \dim(K(T)) \leq \dim(V) < \infty$. The range $R(T)$ is also finite dimensional. In fact, if $\{v_1, ..., v_n\}$ is a basis for V every $w \in R(T)$ has the form $w = T(v) = \sum_i c_i T(v_i)$, so the vectors $T(v_1), ..., T(v_n)$ span $R(T)$. Therefore $\dim(R(T)) \leq n = \dim(V)$.

Now let $\{w_1, ..., w_m\}$ be a basis for $K(T)$. By adjoining additional vectors from V we can obtain a basis $\{w_1, ..., w_m, v_{m+1}, ..., v_{m+k}\}$ for V. Obviously, $m = \dim(K(T))$ and $m + k = \dim(V)$. To get $k = \dim(R(T))$ we show that the vectors $T(v_{m+1}), ..., T(v_{m+k})$ are a basis for $R(T)$. They certainly span $R(T)$ because $w \in R(T) \Rightarrow w = T(v)$ for some $v \in V$, which can be written

$$v = c_1 w_1 + ... + c_m w_m + c_{m+1} v_{m+1} + \ldots + c_{m+k} v_{m+k} \quad (c_j \in \mathbb{K})$$

Since $w_j \in K(T)$ and $T(w_j) = 0_W$ we see that

$$w = T(v) = 0_W + ... + 0_W + \sum_{j=1}^{k} c_{m+j} T(v_{m+j}) \ .$$

so $v \in \mathbb{K}\text{-span}\{T(v_{r+1}, \ldots, T(v_{r+k}\}$.

These vectors are also independent, for if

$$0_W = \sum_{i=1}^{k} c_{m+i} T(v_{m+i}) = T\Big(\sum_{i=1`}^{k} c_{m+i} v_{m+i} \Big)$$

that means $\sum_i c_{m+i} v_{m+i} \in K(T)$ and there are coefficients $c_1, ..., c_m$ such that $\sum_{i=1}^m c_i w_i = \sum_{j=1}^k c_{m+j} v_{m+j}$, or

$$0_V = -c_1 w_1 - ... - c_m w_m + c_{m+1} v_{m+1} + ... + c_{m+k} v_{m+k}$$

Because $\{w_1, ..., v_{m+k}\}$ is a basis for V we must have $c_1 = ... = c_{m+k} = 0$, proving independence of $T(v_{m+1}), ..., T(v_{m+k})$. Thus $\dim(R(T)) = k$ as claimed. $\square$

1.6. Corollary. *Let $T : V \to W$ be a linear operator between finite dimensional vector spaces such that $\dim(V) = \dim(W)$, which certainly holds if $V = W$. Then the following assertions are equivalent:*

(i) *T is one-to-one* (ii) *T is surjective* (iii) *T is bijective.*

Proof: By the Dimension Theorem we have $|K(T)| + |R(T)| = |V|$. If T is one-to-one then $K(T) = (0)$, so by (7) $|R(T)| = |V| = |W|$, Since $R(T) \subseteq W$ the only way that this can happen is to have $R(T) = W$ – i.e. T is surjective. Finally, if T is surjective then $|R(T)| = |W| = |V|$ by hypothesis. Invoking (7) we see that $|K(T)| = 0$, the kernel is trivial, and T is one-to-one.

We just proved that T is one-to-one if and only if T is surjective, so either condition implies T is bijective. $\square$

1.7. Exercise. Explain why a spanning set $\mathfrak{X} = \{v_1, ..., v_n\}$ is a basis for a finite dimensional space if and only the vectors in $\mathfrak{X}$ are independent ($\sum_{i=1}^n c_i v_i = 0_V \Rightarrow c_1 = \ldots = c_n = 0$).

1.8. Proposition. *Let V be a finite dimensional vector space and $\{v_1, ..., v_n\}$ a basis. Select any n vectors $w_1, ..., w_n$ in some other vector space W. Then, there is a unique linear operator $T : V \to W$ such that $T(v_i) = w_i$ for $1 \leq i \leq n$.*

Proof: Uniqueness of T (if it exists) was proved in our initial comments about linear operators. To construct such a T define $T(\sum_{i=1}^{n}\lambda_i v_i) = \sum_{i=1}^{n}\lambda_i w_i$ for all choices of $\lambda_1, ..., \lambda_n \in \mathbb{K}$. This is obviously well-defined since $\{v_i\}$ is a basis, and is easily seen to be a linear operator from $V \to W$. $\square$

One prolific source of linear operators is the correspondence between $n \times m$ matrices A with entries in $\mathbb{K}$ and the linear operators $L_A : \mathbb{K}^m \to \mathbb{K}^n$ determined by matrix multiplication

$$L_A(\mathbf{v}) = A \cdot \mathbf{v} \qquad \Big(\text{matrix product } (n \times m)\cdot(m \times 1) = (n \times 1)\Big) ,$$

if we regard $\mathbf{v} \in \mathbb{K}^m$ as an $m \times 1$ column vector.

1.9. Example. Let $L_A : \mathbb{R}^3 \to \mathbb{R}^4$ (or $\mathbb{C}^3 \to \mathbb{C}^4$, same discussion) be the linear operator associated with the 4×3 matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Describe $\ker(L_A)$ and $\text{range}(L_A)$ by finding explicit basis vectors in these spaces.

Solution: The range $R(L_A)$ of L_A is determined by finding all $\mathbf{y}$ for which there is an $\mathbf{x} \in \mathbb{K}^3$ such that $A\mathbf{x} = \mathbf{y}$. Row reduction of the augmented matrix $[A : \mathbf{y}]$ yields

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & y_1 \\ 1 & 0 & 2 & y_2 \\ 2 & 1 & 1 & y_3 \\ 1 & 1 & 1 & y_4 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & y_1 \\ 0 & -2 & -1 & y_2 - y_1 \\ 0 & -3 & -5 & y_3 - 2y_1 \\ 0 & -1 & -2 & y_4 - y_1 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & y_1 \\ 0 & 1 & 2 & y_1 - y_4 \\ 0 & 2 & 1 & y_1 - y_2 \\ 0 & 3 & 5 & 2y_1 - y_3 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|c} 1 & 2 & 3 & y_1 \\ 0 & 1 & 2 & y_1 - y_4 \\ 0 & 0 & -3 & y_1 - y_2 - 2(y_1 - y_4) \\ 0 & 0 & -1 & 2y_1 - y_3 - 3(y_1 - y_4) \end{array}\right) \to \left(\begin{array}{ccc|c} \boxed{1} & 2 & 3 & y_1 \\ 0 & \boxed{1} & 2 & y_1 - y_4 \\ 0 & 0 & \boxed{1} & y_1 + y_3 - 3y_4 \\ 0 & 0 & 0 & 2y_1 - y_2 + 3y_3 - 7y_4 \end{array}\right)$$

There are no solutions $\mathbf{x} \in \mathbb{K}^3$ unless $\mathbf{y} \in \mathbb{K}^4$ lies the 3-dimensional solution set of the equation

$$2y_1 - y_2 + 3y_3 - 7y_4 = 0.$$

When this constraint is satisfied, backsolving yields exactly one solution for each such $\mathbf{y}$.

Thus $R(L_A)$ is the solution set of equation

$$2y_1 - y_2 + 3y_3 - 7y_4 = 0$$

When this is written as a matrix equation $B\mathbf{y} = \mathbf{0}$ ($B =$ the 1×4 matrix $[2, -1, 3, 7]$), y_2, y_3, y_4 are free variables and then $y_1 = \frac{1}{2}(y_2 - 3y_3 + 7y_4)$, so a typical vector in $R(L_A)$ has the form

$$\mathbf{y} \;=\; \begin{pmatrix} \frac{1}{2}(y_2 - 3y_3 + 7y_4) \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = y_2 \cdot \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + y_3 \cdot \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{pmatrix} + y_4 \cdot \begin{pmatrix} \frac{7}{2} \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

with $y_1, y_2, y_3 \in \mathbb{K}$. The column vectors $\mathbf{u}_2 = \text{col}(1, 2, 0, 0)$, $\mathbf{u}_3 = \text{col}(-3, 0, 2, 0)$, $\mathbf{u}_4 = \text{col}(7, 0, 0, 2)$ obviously span $R(L_A)$ and are easily seen to be linearly independent, so they are a basis for the range, which has dimension $|R(L_A)| = 3$.

The kernel: Now we want to find all solutions $\mathbf{x} \in \mathbb{K}^3$ of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. The same row operations used above transform $[A : \mathbf{0}]$ to:

$$\left(\begin{array}{ccc|c} \boxed{1} & 2 & 3 & 0 \\ 0 & \boxed{1} & 2 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Then $\mathbf{x} = \text{col}(x_1, x_2, x_3)$ has entries $x_1 = x_2 = x_3 = 0$. Therefore

$$K(L_A) = \{\mathbf{x} \in \mathbb{K}^3 : L_A(\mathbf{x}) = A\mathbf{x} = \mathbf{0}\} \text{ is the trivial subspace } \{\mathbf{0}\}$$

and there is no basis to be found.

Note that $|R(L_A)| + |K(L_A)| = 3 + 0 =$ dimension of the *initial space* $\mathbb{K}^3$, while the target space $W = \mathbb{K}^4$ has dimension $= 4$. $\square$

1.10. Example. Let $L_A : \mathbb{R}^4 \to \mathbb{R}^4$ be the linear operator associated with the 4×4 matrix

$$A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 1 & 0 & 2 & -1 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Describe the kernel $K(L_A)$ and range $R(L_A)$ by finding basis vectors.

Solution: The range of L_A is determined by finding all $\mathbf{y}$ such that $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^4$. Row reduction of $[A : \mathbf{y}]$ yields

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 0 & y_1 \\ 1 & 0 & 2 & -1 & y_2 \\ 2 & 1 & 1 & 2 & y_3 \\ 1 & 1 & 1 & 1 & y_4 \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 2 & 3 & 0 & y_1 \\ 0 & -2 & -1 & -1 & y_2 - y_1 \\ 0 & -3 & -5 & 2 & y_3 - 2y_1 \\ 0 & -1 & -2 & 1 & y_4 - y_1 \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 2 & 3 & 0 & y_1 \\ 0 & 1 & 2 & -1 & y_1 - y_4 \\ 0 & 2 & 1 & 1 & y_1 - y_2 \\ 0 & 3 & 5 & -2 & 2y_1 - y_3 \end{array}\right)$$

$$\to \left(\begin{array}{cccc|c} 1 & 2 & 3 & 0 & y_1 \\ 0 & 1 & 2 & -1 & y_1 - y_4 \\ 0 & 0 & -3 & 3 & y_1 - y_2 - 2(y_1 - y_4) \\ 0 & 0 & -1 & 1 & 2y_1 - y_3 - 3(y_1 - y_4) \end{array}\right) \to \left(\begin{array}{cccc|c} \boxed{1} & 2 & 3 & 0 & y_1 \\ 0 & \boxed{1} & 2 & -1 & y_1 - y_4 \\ 0 & 0 & \boxed{1} & -1 & y_1 + y_3 - 3y_4 \\ \hline 0 & 0 & 0 & 0 & 2y_1 - y_2 + 3y_3 - 7y_4 \end{array}\right)$$

There are *no* solutions $\mathbf{x}$ in $\mathbb{R}^4$ unless $\mathbf{y}$ lies the 3-dimensional solution set of the linear equation

$$2y_1 - y_2 + 3y_3 - 7y_4 = 0$$

– i.e. $\mathbf{y}$ is a solution of the matrix equation

$$C\mathbf{y} = \mathbf{0} \qquad \text{where} \qquad C = \big[2\,,\,-1\,,\,3\,,\,-7\big]_{1\times 4}$$

There exist multiple solutions, and then $R(L_A) = \{\mathbf{y} \in \mathbb{R}^4 : C\mathbf{y} = \mathbf{0}\}$ is nontrivial. Multiplying C by $\frac{1}{2}$ puts it in echelon form, so y_2, y_3, y_4 are free variables in solving $C\mathbf{y} = \mathbf{0}$ and $y_1 = \frac{1}{2}(y_2 - 3y_3 + 7y_4)$. Thus a vector $\mathbf{y} \in \mathbb{R}^4$ is in the range $R(L_A) \Leftrightarrow$

$$\mathbf{y} = \begin{pmatrix} \frac{1}{2}(y_2 - 3y_3 + 7y_4) \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = y_2 \cdot \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + y_3 \cdot \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{pmatrix} + y_4 \cdot \begin{pmatrix} \frac{7}{2} \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

with $y_1, y_2, y_3 \in \mathbb{R}^3$. The column vectors $\mathbf{u}_2 = \text{col}(1, 2, 0, 0)$, $\mathbf{u}_3 = \text{col}(-3, 0, 2, 0)$, $\mathbf{u}_4 = \text{col}(7, 0, 0, 2)$, obviously span $R(L_A)$ and are easily seen to be linearly independent, so they are a basis for the range and $\dim(R(L_A)) = 3$. (Hence also $|K(L_A)| = |V| - |R(L_A)| = 1$ by the Dimension Theorem.)

The kernel: $K(L_A)$ can be found by setting $\mathbf{y} = 0$ in the preceding echelon form of $[A : \mathbf{y}]$, which becomes

$$\left(\begin{array}{cccc|c} \boxed{1} & 2 & 3 & 0 & 0 \\ 0 & \boxed{1} & 2 & -1 & 0 \\ 0 & 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Now x_4 is the free variable and

$$\begin{aligned} x_3 &= x_4 \\ x_2 &= -2x_3 + x_4 = -x_4 \\ x_1 &= -2x_2 - 3x_3 = -x_4 \ , \end{aligned}$$

hence,

$$K(L_A) = \mathbb{K} \cdot \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$$

is one dimensional as expected.

Given a vector $\mathbf{y}$ in the range $R(L_A)$ we can find a particular solution $\mathbf{x}_0$ of $A\mathbf{x} = \mathbf{y}$ by setting the free variable $x_4 = 0$. Then

$$\begin{aligned} x_3 &= x_4 + y_1 + y_3 - 3y_4 = y_1 + y_3 - 3y_4, \\ x_2 &= -2x_3 + x_4 + y_1 - y_4 = -y_1 - 2y_3 + 5y_4, \\ x_1 &= -2x_2 - 3x_3 + y_1 = y_1 + (2y_1 + 4y_3 - 10y_4) + (-3y_1 - 3y_3 + 9y_4) \\ &= y_3 - y_4 \end{aligned}$$

and

$$\mathbf{x}_0 = \begin{pmatrix} y_3 - y_4 \\ -y_1 - 2y_3 + 5y_4 \\ y_1 + y_3 - 3y_4 \\ 0 \end{pmatrix}$$

is a particular solution for $A\mathbf{x} = \mathbf{y}$. The full set of solutions is the additive coset $\mathbf{x}_0 + K(L_A)$ of the kernel of L_A. $\square$

1.11. Exercise. If $A \in \mathrm{M}(n \times m, \mathbb{K})$ prove that

1. $\mathrm{Range}(L_A)$ is equal to $\mathrm{Col}(A) = \mathbb{K}\text{-span}\{C_1, \ldots, C_m\}$, the subspace of $\mathbb{K}^n$ spanned by the columns of A.

so

$$\dim(\mathrm{Range}(L_A)i) = \dim\big(\mathrm{Col}(A)\big)$$

II.2. Invariant Subspaces.

If $T : V \to V$ ($V = W$) is a linear operator, a subspace W is T**-invariant** if $T(W) \subseteq W$. Invariant subspaces are important in determining the structure of T, as we shall see. Note that the subspaces (0), $R(T) = \mathrm{range}(T) = T(V)$, $K(T) = \ker(T)$, and V are all T-invariant. Structural analysis of T proceeds initially by searching for **eigenvectors**: nonzero vectors $v \in V$ such that $T(v)$ is a scalar multiple $T(v) = \lambda \cdot v$, for some $\lambda \in \mathbb{K}$. These are precisely the vectors in $\ker(T - \lambda I)$ where $I : V \to V$ is the identity operator on V. Eigenvectors may or may not exist; when they do they have a story to tell.

2.1. Definition. *Fix a linear map $T : V \to V$ and scalar $\lambda \in \mathbb{K}$. The* **λ-eigenspace** *is*

$$E_\lambda(T) = \{v \in V : T(v) = \lambda \cdot v\} = \{v \in V : \big(T - \lambda I\big)(v) = 0\} = \ker(T - \lambda I)$$

We say that $\lambda \in \mathbb{K}$ is an **eigenvalue** *for T if the eigenspace is nontrivial, $E_\lambda(T) \neq (0)$.*
If V is finite dimensional we will eventually see that the number of eigenvalues is $\leq n$ (possibly zero) because it is the set of roots in $\mathbb{K}$ of the "**characteristic polynomial**"

$$p_T(x) = \det(T - xI) \in \mathbb{K}[x] \ ,$$

which has degree $n = \dim_{\mathbb{K}}(V)$. The set of DISTINCT eigenvalues in $\mathbb{K}$ is called the **spectrum of** T and is denoted

$$\begin{aligned} \mathrm{sp}_{\mathbb{K}}(T) &= \{\lambda \in \mathbb{K} : \text{ such that } T(v) = \lambda \cdot v \text{ for some } v \neq 0\} \\ &= \{\lambda \in \mathbb{K} : E_\lambda \neq (0)\} \end{aligned}$$

Depending on the nature of the ground field, $\mathrm{sp}_{\mathbb{K}}(T)$ may be the empty set; it is always nonempty if $\mathbb{K} = \mathbb{C}$, because every nonconstant polynomial has at least one root in $\mathbb{C}$ (Fundamental Theorem of Algebra). The point is that all eigenspaces E_λ are T-invariant subspaces because T and $(T - \lambda I)$ "commute," hence

$$(T - \lambda I)(Tv) = T\big((T - \lambda I)v\big) = 0 \qquad \text{if} \qquad v \in E_\lambda$$

The E_λ are also "essentially disjoint" from each other in the sense that $E_\mu \cap E_\lambda = (0)$, if $\mu \neq \lambda$. (You can't have $\lambda \cdot v = \mu \cdot v$ (or $(\lambda - \mu) \cdot v = 0$) for nonzero v if $\mu \neq \lambda$.) Note that $\ker(T)$ is the eigenspace corresponding to $\lambda = 0$ since

$$E_{\lambda=0} = \{v : (T - 0 \cdot I)(v) = T(v) = 0\} = \ker(T) \ ,$$

and $\lambda = 0$ is an eigenvalue in $\mathrm{sp}_{\mathbb{K}}(T) \Leftrightarrow$ this kernel is nontrivial. When $\lambda = 1$, $E_{\lambda=1}$ is the set of "fixed points" under the action of T.

$$E_{\lambda=1} = \mathrm{Fix}(T) = \{v : T(v) = v\} \qquad \text{(the fixed points in } V\text{)}$$

Decomposition of Operators.

We now show that if $W \subseteq V$ is an **invariant subspace**, so $T(W) \subseteq W$, then T induces linear operators in W and in the quotient space V/W:

1. RESTRICTION: $T|_W : W \to W$ is the restriction of T to W, so $\big(T|_W\big)(w) = T(w)$, for all $w \in W$.
2. QUOTIENT OPERATOR: The operator $\tilde{T} : V/W \to V/W$, sometimes denoted $T_{V/W}$, is induced by the action of T on additive cosets:

   (8) $$T_{V/W}(x + W) = T(x) + W \qquad \text{for all cosets in } V/W \ .$$

   The outcome is determined using a representative x for the coset, but as shown below different representatives yield the same result, so $T_{V/W}$ is well defined.

2.2. Theorem. *Given a linear operator $T : V \to V$ and an invariant subspace W, there is a unique linear operator $\tilde{T} : V/W \to V/W$ that makes the following diagram "commute" in the sense that $\pi \circ T = \tilde{T} \circ \pi$, where $\pi : V \to V/W$ is the quotient map.*

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ \pi \downarrow & & \downarrow \pi \\ V/W & \xrightarrow{\tilde{T}} & V/W \end{array}$$

Note: We have already shown that the quotient map $\pi : V \to V/W$ is a linear operator between vector spaces.

Proof: Existence. The restriction $T|_W$ is clearly a linear operator on W. As for the induced map $\tilde{T}$ on V/W, the fact that $T(W) \subseteq W$ implies

$$T(x+W) = T(x) + T(W) \subseteq T(x) + W \qquad \text{for } x \in V$$

suggests that (8) is the right definition. It automatically insures that $\tilde{T} \circ \pi = \pi \circ T$; the problem is to show the outcome $\tilde{T}(x+W) = T(x) + W$ is independent of the choice of coset representative $x \in V$. For this, suppose $x' + W = x + W$. Then $x' = x' + 0 = x + w_0$ for some $w_0 \in W$ and

$$T(x') + W = T(x + w_0) + W = T(x) + \big(T(w_0) + W\big)$$

Since W is invariant we have $T(w_0) \in W$, and $T(w_0) + W = W$ by Exercise 5.2 of Chapter 1. Hence $T(v') + W = T(v) + W$ and the induced operator in (8) is well-defined. Linearity of $\tilde{T}$ is easily checked: once we know $\tilde{T}$ is well-defined we get

$$\begin{aligned}
\tilde{T}\big((v_1 + W) \oplus (v_2 + W)\big) &= \tilde{T}\big(v_1 + v_2 + W\big) \\
&= T(v_1 + v_2) + W \;=\; T(v_1) + T(v_2) + W \\
&= \big(T(v_1) + W\big) + \big(T(v_2) + W\big) \qquad (\text{since } W + W = W) \\
&= \tilde{T}(v_1 + W) \oplus \tilde{T}(v_2 + W)
\end{aligned}$$

and similarly

$$\tilde{T}\big(\lambda \odot (v + W)\big) = \tilde{T}(\lambda \cdot v + W) = \lambda \odot \tilde{T}(v + W)$$

Uniqeuness. If $\tilde{T}_1$, $\tilde{T}_2$ both satisfy the commutation relation $\tilde{T}_i \circ \pi = \pi \circ T_i$, then

$$\tilde{T}_1(v + W) = \tilde{T}_1 \circ \pi(v) = \pi(T(v)) = \tilde{T}_2(\pi(v)) = \tilde{T}_2(v + W)$$

so $\tilde{T}_1 = \tilde{T}_2$ on V/W. $\square$

A Look Ahead. If $T : V \to V$ is a linear operator on a finite dimensional space, we will explain in Section II.4 how a matrix $[T]_{\mathfrak{X}}$ is associated with T once a basis $\mathfrak{X}$ in V has been specified. If W is a T-invariant subspace we will see that much of the structural information about T resides in the induced operators $T|_W$ and $T_{V/W}$, and that in some sense (to be made precise) T is assembled by "joining together" these smaller pieces. This is a big help in trying to understand the action of T on V, but it does depend on being able to find invariant subspaces – the more the better! To illustrate: if a basis $\mathfrak{X} = \{w_1, \ldots, w_m\}$ in W is augmented to get a basis for V,

$$\mathfrak{Z} = \{w_1, \ldots, w_m, v_{m+1}, \ldots, v_{m+k}\} \qquad (m + k = n = \dim(V))$$

we have seen that the image vectors $\bar{v}_{m+i} = \pi(v_{m+i})$ are a basis $\mathfrak{Y}$ in the quotient space V/W. In Section II.4 we will show that the matrix $[T]_{\mathfrak{Z}}$ assumes a special "**block-upper triangular form**" with respect to such a basis.

$$[T]_{\mathfrak{Z}} = \begin{pmatrix} \boxed{\text{A}}_{m \times m} & \boxed{\;*\;}_{m \times k} \\ \mathbf{0}_{k \times m} & \boxed{\text{B}}_{k \times k} \end{pmatrix}$$

where $A = [T|_W]_{\mathfrak{X}}$ and $B = [T_{V/W}]_{\mathfrak{Y}}$. Clearly, much of the information about T is encoded in the two diagonal blocks A, B; but some information is lost in passing from T to $T|_W$ and $T_{V/W}$ – the "cross-terms" in the upper right block $\boxed{\;*\;}$ cannot be determined if we only know the two induced operators. Additional information is needed to piece them together to recover T, and there may be more than one operator T yielding a particular pair of induced operators $\big(T|_W, T_{V/W}\big)$. $\square$

Isomorphisms of Vector Spaces. A linear map $T : V \to W$ is an **isomorphism** between vector spaces if it is a bijection. Since T is a bijection there is a well-defined "set-theoretic" inverse map in the opposite direction $T^{-1} : W \to V$

$$T^{-1}(w) = (\textit{the unique } v \in V \textit{ such that } T(v) = w)$$

for any $w \in W$. In general it might not be easy to describe the inverse of a bijection $f : X \to Y$ between two point sets in closed form

$$f^{-1}(y) = (\textit{some explicit formula})$$

(Try finding $x = f^{-1}(y)$ where $f : \mathbb{R} \to \mathbb{R}$ is $y = f(x) = x^3 + x + 1$, which is a bijection because $df/dx > 0$ for all x.) But the inverse of a linear map is automatically linear, if it exists. We write $V \cong W$ if there is an isomorphism between them.

2.3. Exercise. Suppose $T : V \to W$ is linear and a bijection. Prove that the set-theoretic inverse map $T^{-1} : W \to V$ must be linear, so

$$T^{-1}(\lambda w) = \lambda{\cdot}T^{-1}(w) \qquad \text{and} \qquad T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2)$$

Thus T and T^{-1} are both *isomorphisms* between V and W.
Hint: T^{-1} reverses the action of T and vice-versa, so $T \circ T^{-1} = \mathrm{id}_W$ and $T^{-1} \circ T = \mathrm{id}_V$.

We now observe that an isomorphisms between vector spaces V and W identifies important features of V with those of W. It maps

$$\left\{\begin{array}{c}\text{independent sets}\\ \text{spanning sets}\\ \text{bases}\end{array}\right. \quad \text{in } V \quad \longrightarrow \quad \left\{\begin{array}{c}\text{independent sets}\\ \text{spanning sets}\\ \text{bases}\end{array}\right. \quad \text{in } W$$

To illustrate, if $\{v_1, ..., v_n\}$ are independent in V, then

$$0_W = \sum_{i=1}^{n} c_i T(v_i) = T\Big(\sum_i c_i v_i\Big) \quad \Rightarrow \quad 0_V = \sum_{i=1}^{n} c_i v_i \text{ in } V \quad \Rightarrow \quad c_1 = ... = c_n = 0 \text{ in } \mathbb{K}$$

because $T(0_V) = 0_W$ and T is one-to-one. Thus the vectors $\{T(v_1), ..., T(v_n)\}$ are independent in W. Similar arguments yield the other two assertions.

2.4. Exercise. If $T : V \to W$ is an isomorphism of vector spaces, verify that:

1. $\mathbb{K}\text{-span}\{v_i\} = V \Rightarrow \mathbb{K}\text{-span}\{T(v_i)\} = W$;
2. $\{v_i\}$ is a basis in $V \Rightarrow \{T(v_i)\}$ is a basis in W.

In particular, isomorphic vector spaces V, W are either both infinite dimensional, or both finite dimensional with $\dim_{\mathbb{K}}(V) = \dim_{\mathbb{K}}(W)$.

The following result which relates linear operators, quotient spaces, and isomorphisms will be cited often in analyzing the structure of linear operators. It is even valid for infinite dimensional spaces.

2.5. Theorem (First Isomorphism Theorem). *Let $T : V \to R(T) \subseteq W$ be a linear map with range $R(T)$. If $K(T) = \ker(T)$, there is a unique bijective linear map $\tilde{T} : V/K(T) \to R(T)$ that makes the following diagram "commute" ($\tilde{T} \circ \pi = T$).*

$$\begin{array}{ccc} V & \xrightarrow{T} & R(T) \subseteq W \\ \pi \downarrow & \nearrow_{\tilde{T}} & \\ V/K(T) & & \end{array}$$

where $\pi : V \to V/K$ is the quotient map. Furthermore $R(\tilde{T}) = R(T)$,
Hints: Try defining $\tilde{T}(v + K(T)) = T(v)$. Your first task is to show that the outcome is independent of the particular coset representative – i.e. $v' + K(T) = v + K(T) \Rightarrow T(v') = T(v)$, so $\tilde{T}$ is well-defined. Next show $\tilde{T}$ is linear, referring to the operations $\oplus$ and $\odot$ in the quotient space $V/K(T)$. The commutation property $\tilde{T} \circ \pi = T$ is built into the definition of $\tilde{T}$. For uniqueness, you must show that if $S : V/K(T) \to W$ is any other linear map such that $S \circ \pi = T$, then $S = \tilde{T}$; this is trivial once you clearly understand the question. $\square$

Note that range($\tilde{T}$) = range(T) because the quotient map $\pi : V \to V/K$ is surjective: $w \in R(\tilde{T}) \Leftrightarrow$ there is a coset $v + K(T)$ such that $\tilde{T}(v + K(T)) = w$; but then $T(v) = w$ and $w \in R(T)$. Since $\tilde{T}$ is an isomorphism between $V/K(T)$ and $R(T)$, $\dim(V/K(T)) = \dim(V) - \dim(K(T))$ is equal to $\dim(R(T))$.

II.3. (Internal) Direct Sum of Vector Spaces.

A vector space V is an **(internal) direct sum** of subspaces V_1, ... ,V_n, indicated by writing $V = V_1 \oplus \ldots \oplus V_n$, if

1. The linear span $\sum_{i=1}^n V_i = \{\sum_{i=1}^n v_i : v_i \in V\}$ is all of V;
2. Every $v \in V$ has a *unique* representation as a sum $v = \sum_{i=1}^n v_i$ with $v_i \in V_i$.

Once we know that the V_i span V, condition (2.) is equivalent to saying

$$2^* \quad \sum_{i=1}^n v_i = 0 \quad \text{with } v_i \in V_i \quad \to \quad v_i = 0 \text{ for all } i.$$

In fact, if a vector v has two different representations $v = \sum_i v_i = \sum v_i'$ then $0 = \sum_{i=1}^n v_i''$ with $v_i'' = (v_i' - v_i) \in V_i$. Then ($2^*$.) implies $v_i'' = 0$ and $v_i' = v_i$ for all i. Conversely, if we can write $0 = \sum w_i$ with $w_i \in V_i$ not all zero, then the representation of a vector as $v = \sum_i v_i$ $(v_i \in V_i)$ cannot be unique, since we could also write

$$0 = \sum_i w_i \quad \text{ with } w_i \neq 0 \text{ for some } i\ ,$$

and then $v = v + 0 = \sum_i (v_i + w_i)$ in which $v_i + w_i \in V_i$ is $\neq v_i$. Thus the condition (2.) is equivalent to (2^*.)

3.1. Example. We note the following examples of direct sum decompositions.

1. $\mathbb{K}^n = V \oplus W$ where

   $$V = \{(x_1, x_2, 0, ..., 0) : x_1, x_2 \in \mathbb{K}\} \text{ and } W = \{(0, 0, x_3, ..., x_n) : x_k \in \mathbb{K}\};$$

   More generally, in an obvious sense we have $\mathbb{K}^{m+n} \cong \mathbb{K}^m \oplus \mathbb{K}^n$.

2. The space of polynomials $\mathbb{K}[x] = V \oplus W$ is a direct sum of the subspaces
   - Even Polynomials: $V = \{ \sum_{i=0}^\infty a_i x^i : a_i = 0 \text{ for odd indices}\}$
   - Odd Polynomials: $W = \{ \sum_{i=0}^\infty a_i x^i : a_i = 0 \text{ for even indices } \}$

3. For $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, matrix space $\mathrm{M}(n, \mathbb{K})$ is a direct sum $\mathcal{A} \oplus \mathcal{S}$ of
   - Antisymmetric Matrices: $\mathcal{A} = \{A : A^{\mathrm{t}} = -A\}$
   - Symmetric Matrices: $\mathcal{S} = \{A : A^{\mathrm{t}} = A\}$.

In fact, since $(A^{\mathrm{t}})^{\mathrm{t}} = A$ we can write any matrix as

$$A = \frac{1}{2}(A^{\mathrm{t}} + A) + \frac{1}{2}(A^{\mathrm{t}} - A)$$

The first term is symmetric and the second antisymmetric, so $\mathrm{M}(n, \mathbb{K}) = \mathcal{A} + \mathcal{S}$, (linear span).

If $B \in \mathcal{A} \cap \mathcal{S}$ then $B^{\mathrm{t}} = -B$ and also $B^{\mathrm{t}} = B$, hence $B = -B$ and $B = \mathbf{0}$ (the zero matrix). Thus $\mathcal{A} \cap \mathcal{S} = (0)$ and Exercise 3.2 (below) implies that $\mathrm{M}(n, \mathbb{K}) = \mathcal{A} \oplus \mathcal{S}$.

Note: This actually works for any field $\mathbb{K}$ in which $2 = 1 + 1 \neq 0$ because the "$\frac{1}{2}$" in the formulas involves division by 2. In particular it works for the finite fields $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ except for $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$, in which $[1] \oplus [1] = [1+1] = [2] = [0]$ $\square$

If subspaces $V_1, \ldots, V_n$ span V it can be tricky to verify that V is a direct sum when $n \geq 3$, but if there are just two summands V_1 and V_2 (the case most often encountered) there is a simple and extremely useful criterion for deciding whether $V = V_1 \oplus V_2$.

3.2. Exercise. If E, F are subspaces of V show that V is the direct sum $E \oplus F$ if and only if

1. They span V: $E + F = \{a + b : a \in E, b \in F\}$ is all of V;
2. Trivial intersection: $E \cap F = \{0\}$.

It is important to note that this is NOT true when $n \geq 3$. If $\sum_{i=1}^{n} V_i = V$ and the spaces are only "pairwise disjoint,"

$$V_i \cap V_j = (0) \text{ for } i \neq j,$$

this is not enough to insure that V is a direct sum of the given subspaces (see the following exercise).

3.3. Exercise. Find three distinct 1-dimensional subspaces V_i in the two dimensional space $\mathbb{R}^2$ such that

1. $V_i \cap V_j = (0)$ for $i \neq j$;
2. $\sum_{i=1}^{3} V_i = \mathbb{R}^2$

Explain why $\mathbb{R}^2$ is not a *direct* sum $V_1 \oplus V_2 \oplus V_3$ of these subspaces.

3.4. Exercise. If $V = V_1 \oplus \ldots \oplus V_n$ and V is finite dimensional, we have seen that each V_i must be finite-dimensional with $\dim(V_i) \leq \dim(V)$.

1. Given bases $\mathfrak{X}_1 \subseteq V_1$, ... , $\mathfrak{X}_n \subseteq V_n$, explain how to create a basis for all of V;
2. Prove

(9) DIMENSION FORMULA FOR SUMS: $$\dim_{\mathbb{K}}(V_1 \oplus \ldots \oplus V_n) = \sum_{i=1}^{n} \dim_{\mathbb{K}}(V_i)$$

Direct sum decompositions play a large role in understanding the structure of linear operators. Suppose $T : V \to V$ and $V = V_1 \oplus V_2$, and that *both* subspaces are T-invariant. We get restricted operators $T_1 = T|_{V_1} : V_1 \to V_1$ and $T_2 = T|_{V_2} : V_2 \to V_2$, but because both subspaces are T-invariant we can fully reconstruct the original operator T in V from its "components" T_1 and T_2. In fact, every $v \in V$ has a unique decomposition $v = v_1 + v_2$ ($v_i \in V_i$) and then

$$T(v) = T(v_1) + T(v_2) = T_1(v_1) + T_2(v_2)$$

We often indicate this decomposition by writing $T = T_1 \oplus T_2$.

This does not work if only one subspace is invariant. But when both are invariant and we take bases $\mathfrak{X}_1 = \{v_1, \ldots, v_m\}$, $\mathfrak{X}_2 = \{v_{m+1}, \ldots, v_{m+k}\}$ for V_1, V_2, we will soon see that the combined basis $\mathfrak{Y} = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_{m+k}\}$ for all of V yields a matriix of particularly simple "**block-diagonal**" form

$$[T]_{\mathfrak{Y}} = \begin{pmatrix} \boxed{\mathrm{A}}_{m\times m} & \mathbf{0}_{\,m\times k} \\ \mathbf{0}_{\,k\times m} & \boxed{\mathrm{B}}_{k\times k} \end{pmatrix}$$

where A, B are the matrices of T_1, T_2 with respect to the bases $\mathfrak{X}_1 \subseteq V_1$ and $\mathfrak{X}_2 \subseteq V_2$.

Projections and Direct Sums. If $V = V_1 \oplus \ldots \oplus V_n$ then for each i there is a natural **projection operator** $P_i : V \to V_i \subseteq V$, the "*projection of V onto V_i along the complementary subspace $\bigoplus_{j\neq i} V_j$.*" By definition we have

$$P_i(v) = v_i \text{ if } v = \sum_{j=1}^{n} v_j \text{ is the unique decomposition with } v_j \in V_j. \tag{10}$$

A number of properties of these projection operators are easily verified.

3.5. Exercise. Show that the projections P_i associated with a direct sum decomposition $V = V_1 \oplus \ldots \oplus V_n$ have the following properties.

1. Linearity: Each $P_i : V \to V$ is a linear operator;
2. Idempotent Property: $P_i^2 = P_i \circ P_i = P_i$ for all i;
3. $P_i \circ P_j = 0$ if $i \neq j$;
4. range$(P_i) = V_i$ and $\ker(P_i)$ is the linear span $\sum_{j\neq i} V_j$;
5. $P_1 + \ldots + P_n = I$ (identity operator on V).

If we represent vectors $v \in V$ as ordered n-tuples $(v_1, ..., v_n)$ in the Cartesian product set $V_1 \times \ldots \times V_n$, the i^{th} projection takes the form

$$P_i(v_1, ..., v_n) = (0, ..., 0, v_i, 0, ..., 0) \in V_i \subseteq V.$$

Don't be misled by this notation into thinking that we are speaking of orthogonal projections (onto orthogonal subspaces in $\mathbb{R}^n$). The following example and exercises illustrate what's really happening.

Note: A vector space must be equipped with additional structure such as an *inner product* if we want to speak of "orthogonality of vectors," or their "lengths." Such notions are meaningless in an unadorned vector space. Nevertheless, inner product spaces are important and will be fully discussed in Chapter VI.

3.6. Example. The plane $\mathbb{R}^2$ is a direct sum of the subspaces $V_1 = \mathbb{R}\mathbf{e}_1$ and $V_2 = \mathbb{R}(\mathbf{e}_1 + \mathbf{e}_2)$, where $\mathfrak{X} = \{\mathbf{e}_2, \mathbf{e}_2\}$ are the standard basis vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ in $\mathbb{R}^2$. The maps

- P_1 projecting V onto V_1 along V_2,
- P_2 projecting V onto V_2 along V_1

are *oblique projections*, not the familiar *orthogonal projections* sending $\mathbf{x} = (x_1, x_2)$ to $(x_1, 0)$ and to $(0, x_2)$ respectively, see Figure 2.1. Find an explicit formula for $P_i(v_1, v_2)$, for arbitrary pairs (v_1, v_2) in $\mathbb{R}^2$.

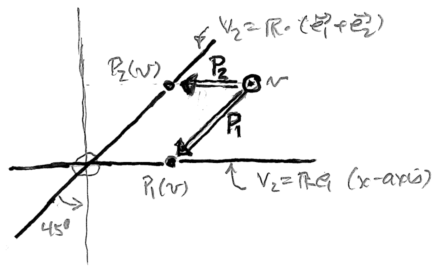

Figure 2.2. Projections P_1, P_2 determined by a direct sum decomposition $V = V_1 \oplus V_2$. Here $V = \mathbb{R}^2$, $V_1 = \mathbb{R}\mathbf{e}_1$, $V_2 = \mathbb{R}(\mathbf{e}_1 + \mathbf{e}_2)$; P_1 projects vectors $v \in V$ onto V_1 along V_2, and likewise for P_2.

Discussion: To calculate these projections we must write an arbitrary vector $\mathbf{v} = (v_1, v_2) = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$ in the form $a + b \in V_1 \oplus V_2$ where $V_1 = \mathbb{R}\mathbf{e}_1$ and $V_2 = \mathbb{R}(\mathbf{e}_1 + \mathbf{e}_2)$. The vectors $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$

$$\mathbf{f}_1 = \mathbf{e}_1 \qquad \text{and} \qquad \mathbf{f}_2 = \mathbf{e}_1 + \mathbf{e}_2 \tag{11}$$

that determine the 1-dimensional spaces V_1, V_2 are easily seen to be a new basis for $\mathbb{R}^2$. If $v = c_1\mathbf{f}_1 + c_2\mathbf{f}_2$ in the new basis, the action of the projections P_1, P_2 can be written immediately based on the definitions:

$$\begin{aligned} P_1(c_1\mathbf{f}_1 + c_2\mathbf{f}_2) &= c_1\mathbf{f}_1 \qquad \big(= c_1\mathbf{e}_1 \big) \\ P_2(c_1\mathbf{f}_1 + c_2\mathbf{f}_2) &= c_2\mathbf{f}_2 \qquad \big(= c_2(\mathbf{e}_1 + \mathbf{e}_2) \big) \end{aligned} \tag{12}$$

Now $v = (v_1, v_2)$ is $v_1\mathbf{e}_1 + v_2\mathbf{e}_2$ in terms of the standard basis $\mathfrak{X}$ in $\mathbb{R}^2$ and we want to describe the outcomes $P_1(v), P_2(v)$ in terms of the same basis.

First observe that the action of P_2 is known as soon as we know the action i of P_1: by the Parallelogram Law for vector addition (see Figure 2.2) we have $P_1 + P_2 = I$, so

$$P_2(v) = (I - P_1)(v) = v - P_1(v) \qquad \text{for all } v \in \mathbb{R}^2$$

Second, the projections P_i are linear so their action is known once we determine the images $P_i(\mathbf{e}_k)$ of the basis vectors $\mathbf{e}_k$ because

$$P_i(v) = P_1(v_1, v_2) = P_i\big(v_1\mathbf{e}_1 + v_2\mathbf{e}_2\big) \;=\; v_1 \cdot P_i(\mathbf{e}_1) + v_2 \cdot P_i(\mathbf{e}_2) \qquad (v_1, v_2 \in \mathbb{R})$$

The last step is to use the linear equations (11) to write the standard basis vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$ in terms of the new basis $\{\mathbf{f}_1, \mathbf{f}_2\}$; then the action of P_1 in the standard basis is easily evaluated by applying (12). From (11) we get

$$\mathbf{e}_1 = \mathbf{f}_1 \qquad \text{and} \qquad \mathbf{e}_2 = \mathbf{f}_2 - \mathbf{f}_1$$

and then from (12),

$$\begin{aligned} P_1(\mathbf{e}_1) &= P_1(\mathbf{f}_1) \;=\; \mathbf{f}_1 \;=\; \mathbf{e}_1 \\ P_1(\mathbf{e}_2) &= P_1(\mathbf{f}_2 - \mathbf{f}_1) \;=\; -\mathbf{f}_1 \;=\; -\mathbf{e}_1 \end{aligned}$$

and similarly

$$\begin{aligned} P_2(\mathbf{e}_1) &= P_2(\mathbf{f}_1) \;=\; \mathbf{0} \\ P_2(\mathbf{e}_2) &= P_2(\mathbf{f}_2 - \mathbf{f}_1) \;=\; \mathbf{f}_2 \;=\; \mathbf{e}_1 + \mathbf{e}_2 \end{aligned}$$

The projections P_i can now be written in terms of the Cartesian coordinates in $\mathbb{R}^2$ as

$$\begin{aligned} P_1(v_1, v_2) &= v_1 P_1(\mathbf{e}_1) + v_2 P_1(\mathbf{e}_2) \\ &= v_1 \cdot \mathbf{e}_1 - v_2 \cdot \mathbf{e}_1 \;=\; (v_1 - v_2, 0) \\ P_2(v_1, v_2) &= v_1 P_2(\mathbf{e}_1) + v_2 P_2(\mathbf{e}_2) \\ &= v_1 \cdot \mathbf{0} + v_2 \cdot (\mathbf{e}_1 + \mathbf{e}_2) \;=\; (v_2, v_2) \end{aligned}$$

It is interesting to calculate P_1^2, P_2^2 and $P_1 \circ P_2$ using the preceding formulas to verify the properties listed in Exercise 3.5 $\square$

The "idempotent property" $P^2 = P$ for a linear operator is characteristic of projections associated with a direct sum decomposition $V = V_1 \oplus V_2$. We have already seen that if $P, Q = (I - P)$ are the projections associated with such a decomposition, then

(i) $P^2 = P$ and $Q^2 = Q$ $\quad$ (ii) $PQ = QP = 0$ $\quad$ (iii) $P + Q = I$ (identity operator)

But the converse is also true.

3.7. Proposition. *If $P : V \to V$ is a linear operator such that $P^2 = P$, then V is a direct sum $V = R(P) \oplus K(P)$ and P is the projection of V onto the range $R(P)$, along the kernel $K(P)$. The operator $Q = I - P$ is also idempotent, with*

$$R(Q) = R(I - P) = K(P) \qquad \textit{and} \qquad K(Q) = K(I - P) = R(P) \;, \tag{13}$$

and projects V onto $K(P)$ along $R(P)$.

Proof: First observe that $Q = (I - P)$ is also idempotent since

$$(I - P)^2 = I - 2P + P^2 = I - 2P + P = (I - P) \;.$$

Next, note that

$$Q(v) = (I - P)v = 0 \;\Leftrightarrow\; P(v) = v \;\Leftrightarrow\; v \in R(P) \,.$$

[Implication ($\Rightarrow$) in the last step is obvious. Conversely, if $v \in R(P)$ with $v = P(w)$, then $P(v) = P^2(w) = P(w) = v$, proving ($\Leftarrow$).] Thus

1. $K(Q) = K(I - P)$ is equal to $R(P)$
2. $R(Q) = R(I - P)$ is equal to $K(P)$,

proving (13).

Obviously $P + Q = I$ because $v = P(v) + (I - P)(v)$ implies $P(v) \in R(P)$, while $(I - P)(v) \in R(Q) = K(P)$ by (13); thus the span $R(P) + K(P) = R(P) + R(Q)$ is all of V. Furthermore $K(P) \cap R(P) = (0)$, for if v is in the intersection we have $v \in K(P) \Rightarrow P(v) = 0$. But we also have $v \in R(P)$, so $v = P(w)$ for some w, and then

$$0 = P(v) = P^2(w) = P(w) = v \;.$$

By Exercise 3.2 we conclude that $V = R(P) \oplus K(P)$.

Finally, let $\tilde{P}$ be the projection onto $R(P)$ along $K(P)$ associated with this decomposition; we claim that $\tilde{P} = P$. By definition $\tilde{P}$ maps $v = r + k \in R(P) \oplus K(P)$ to r; that, however, is exactly what our original operator P does:

$$P(r + k) = P(r) + P(k) = r + 0 = r \;.$$

Therefore $\tilde{P} = P$ as operators on V. $\square$

Direct Sums and Eigenspaces. Let $T : V \to V$ be a linear operator on a finite dimensional space V. As above, the spectrum of T is the set of *distinct* eigenvalues $\mathrm{sp}(T) = \{\lambda \in \mathbb{K} : E_\lambda \neq 0\}$, where E_λ is the (nontrivial) λ-eigenspace

$$E_\lambda = \{v \in V : T(v) = \lambda{\cdot}v\} = \ker(T - \lambda I) \qquad\qquad (I = \mathrm{id}_V)$$

3.8. Definition. *A linear operator $T : V \to V$ is* **diagonalizable** *if V is the direct sum of the nontrivial eigenspaces,*

$$V = \bigoplus_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$$

We will see below that this happens if and only if V has a basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ of eigenvectors,

$$T(\mathbf{f}_i) \;=\; \mu_i{\cdot}\mathbf{f}_i \qquad\qquad \text{for some } \mu_i \in \mathbb{K}$$

for $1 \leq i \leq n$.

Our next result shows that T is actually diagonalizable if we only know that the eigenspaces *span* V, with

$$V \;=\; \sum_{\lambda \in \mathrm{sp}(T)} E_\lambda(T) \;=\; \mathbb{K}\text{-span}\{E_\lambda : \lambda \in \mathrm{sp}_{\mathbb{K}}(T)\}$$

(a property much easier to verify).

3.9. Proposition. *If $T : V \to V$ is a linear operator on a finite dimensional space, let W be the span $\sum_{\lambda\in\mathrm{sp}_{\mathbb{K}}(T)} E_\lambda(T)$ of the eigenspaces. This space is T-invariant and is in fact a direct sum $W = \bigoplus_\lambda E_\lambda$ of the eigenspaces.*

Proof: Since each E_λ is invariant their span W is also T-invariant. The E_λ span W by hypothesis, so each $w \in W$ has some decomposition $w = \sum_\lambda w_\lambda$ with $w_\lambda \in E_\lambda$. For uniqueness of this decomposition it suffices to show that

$$0 = \sum_{\lambda \in \mathrm{sp}(T)} w_\lambda \text{ with } w_\lambda \in E_\lambda \quad\Rightarrow\quad \text{each } w_\lambda = 0\ .$$

The operators $T, (T - \lambda I)$, and $(T - \mu I)$ commute for all $\mu, \lambda \in \mathbb{K}$ since the identity element I and its scalar multiples commute with everybody. Let us fix an eigenvalue λ_0; we will show $w_{\lambda_0} = 0$. With this λ_0 in mind we define the product

$$A \;=\; \prod_{\lambda\neq\lambda_0, \lambda\in\mathrm{sp}(T)} (T - \lambda I)\ .$$

Then

$$0 = A(0) = A\Big(\sum_\lambda w_\lambda\Big) = \sum_\lambda A(w_\lambda) = A(w_{\lambda_0}) + \sum_{\lambda\neq\lambda_0} A(w_\lambda)$$

If $\lambda \neq \lambda_0$ we have

$$A(w_\lambda) = \left(\prod_{\mu\neq\lambda_0} (T - \mu I)\right) w_\lambda = \left(\prod_{\mu\neq\lambda_0,\lambda} (T - \mu I)\right) \cdot (T - \lambda I)\, w_\lambda = 0$$

because $w_\lambda \in E_\lambda$. On the other hand, by writing $(T - \lambda_0 I) + (\lambda_0 - \mu)I$ we find that

$$A(w_{\lambda_0}) = \left(\prod_{\mu\neq\lambda_0} (T - \mu I)\right) w_{\lambda_0} \;=\; \left(\prod_{\mu\neq\lambda_0} (T - \lambda_0 I) + (\lambda_0 - \mu)I\right) w_{\lambda_0}$$

When we expand this product of sums, every term but one includes a factor $(T - \lambda_0 I)$ that kills w_{λ_0}:

$$(Term) = (product\ of\ operators) \cdot (T - \lambda_0 I)\, w_{\lambda_0} \;=\; 0$$

The one exception is the product

$$\Big(\prod_{\mu \neq \lambda_0} (\lambda_0 - \mu)\Big) \cdot w_{\lambda_0}.$$

The scalar out front cannot be zero because each $\mu \neq \lambda_0$, so

$$A(w_{\lambda_0}) = \prod_{\mu \neq \lambda_0} (\lambda_0 - \mu) \cdot w_{\lambda_0} \neq 0$$

But we already observed that

$$0 = A\Big(\sum_{\lambda} w_\lambda\Big) = 0 + A(w_{\lambda_0})$$

so we get a contradiction unless $w_{\lambda_0} = 0$. Thus each term in $\sum_\lambda w_\lambda$ is zero and W is the *direct* sum of the eigenspaces E_λ. $\square$

If $W \subsetneqq V$ this result tells us nothing about the behavior of T off of the subspace W, but if we list the distinct eigenvalues as $\lambda_1, \ldots, \lambda_r$ we can construct a basis for W that runs first through E_{λ_1}, then through E_{λ_2}, etc to get a basis for W,

$$\mathfrak{X} = \{f_1^{(1)}, \ldots, f_{d_1}^{(1)}; f_1^{(2)}, \ldots, f_{d_2}^{(2)}; \ldots; f_1^{(r)}, \ldots, f_{d_r}^{(r)}\}$$

where $d_i = \dim(E_{\lambda_i})$ and $\sum_i d_i = m = \dim(W)$. The corresponding matrix describing $T|_W$ is *diagonal*, so $T|_W$ is a diagonalizable operator on W even if T is not diagonalizable on all of V.

$$[T|_W]_{\mathfrak{X},\mathfrak{X}} \;=\; \begin{pmatrix} \lambda_1 & & & & & & 0 \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_r & & \\ & & & & & \ddots & \\ 0 & & & & & & \lambda_r \end{pmatrix} \tag{14}$$

where $\mathrm{sp}(T) = \{\lambda_1, ..., \lambda_r\}$.

3.10. Exercise. If a linear operator $T : V \to V$ acts on a finite dimensional space, prove that the following statements are equivalent.

1. T is diagonalizable: $V = \bigoplus_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$
2. There is a basis $f_1, \ldots, f_n$ for V such that each f_i is an eigenvector, with $Tf_i = \mu_i f_i$ for some $\mu_i \in \mathbb{K}$.

II.4. Representing Linear Operators as Matrices.

Let $T : V \rightarrow W$ be a linear operator between finite dimensional vector spaces with $\dim(V) = m$, $\dim(W) = n$. An **ordered basis** in V is an ordered list $\mathfrak{X} = \{e_1, \dots, e_n\}$ of vectors that are independent and span V; let $\mathfrak{Y} = \{f_1, ..., f_n\}$ be an ordered basis for the target space W.

The behavior of a linear map $T : V \rightarrow W$ is completely determined by what it does to the basis vectors in V because every $v = \sum_{i=1}^m c_i e_i$ (uniquely) and

$$T\big(\sum_{i=1} c_i e_i\big) = \sum_{i=1}^{n} c_i T(e_i)$$

Each image $T(e_i)$ can be expressed uniquely as a linear combination of vectors in the $\mathfrak{Y}$ basis,

$$T(e_i) = \sum_{j=1}^{n} t_{ji} f_j \qquad \text{for } \ 1 \leq i \leq m$$

yielding a system of $m = \dim(V)$ vector equations that tell us how to rewrite vectors in the $\mathfrak{X}$-basis in terms of vectors in the $\mathfrak{Y}$-basis

$$\begin{aligned} T(e_1) \quad &= \quad b_{11} f_1 + ... + b_{1n} f_n \\ &\vdots \\ T(e_m) \quad &= \quad b_{m1} f_1 + ... + b_{mn} f_n \end{aligned} \tag{15}$$

We *define* the **matrix of T with respect to the bases $\mathfrak{X}$, $\mathfrak{Y}$** to be the $n \times m$ matrix

$$[T]_{\mathfrak{Y}\mathfrak{X}} = [t_{ij}] \quad \text{(the } \textit{transpose} \text{ of the array of coefficients } B = [b_{ij}] \text{ in (15))}$$

Since $(B^{\mathrm{t}})_{k\ell} = B_{\ell,k}$ that means $t_{ij} = b_{ji}$; to put it differently, the entries t_{ij} in $[T]_{\mathfrak{Y}\mathfrak{X}}$ satisfy the following identities derived from (15)

$$T(e_i) = \sum_{j=1}^{n} t_{ji} f_j \qquad \text{or } 1 \leq i \leq m \tag{16}$$

Note carefully: the basis vector f_j in (16) is paired with t_{ji} and not t_{ij}.

The matrix description of $T : V \rightarrow W$ changes if we take different bases; nevertheless, the same operator T (which has a coordinate-independent existence) underlies all of these descriptions. One objective in analyzing T is to find bases that yield the simplest matrix descriptions. When $V = W$ the best possible outcome is of course a basis that diagonalizes T as in (14), but alas, not all operators are diagonalizable.

Another issue worth considering is the following: If T is the identity operator $I = \mathrm{id}$ on a vector space V, and we compute $[\mathrm{id}]_{\mathfrak{X}\mathfrak{X}}$, the outcome is the same for all bases $\mathfrak{X}$,

$$[\mathrm{id}]_{\mathfrak{X}\mathfrak{X}} = I_{n\times n} \qquad \text{(the } n \times n \text{ identity matrix)}$$

But there is no reason why we couldn't take different bases in the initial and final spaces (even if they are the same space), regarding $T = \mathrm{id}_V$ as a map from $(V, \mathfrak{X})$ to $(V, \mathfrak{Y})$. Then there are some surprises when you compute $[T]_{\mathfrak{X}\mathfrak{Y}}$.

4.1. Exercise. Let V be 2-dimensional coordinate space $\mathbb{R}^2$. Let $I : V \rightarrow V$ be identity map $I = \mathrm{id}_V$, but take different bases $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\}$ and $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ in the initial and final spaces. Letting $\mathbf{e}_1, \mathbf{e}_2$ be the standard basis vectors in $\mathbb{R}^2$ and $\mathbf{f}_1 = (1, 0), \mathbf{f}_2 = (1, 2)$, compute the matrices

$$(i) \ \ [I]_{\mathfrak{X}\mathfrak{X}} \qquad (ii) \ \ [I]_{\mathfrak{Y}\mathfrak{X}} \qquad (iii) \ \ [I]_{\mathfrak{Y}\mathfrak{Y}} \qquad (iv) \ \ [I]_{\mathfrak{X}\mathfrak{Y}}$$

4.2. Exercise. If V, W are finite dimensional and $T : V \to W$ is linear, prove that there are always bases $\mathfrak{X}, \mathfrak{Y}$ and $\mathfrak{X}', \mathfrak{Y}'$ in V, W such that

$$(i)\;\; [I]_{\mathfrak{Y}\mathfrak{X}} = \begin{pmatrix} 0 & 0 \\ 0 & I_{r\times r} \end{pmatrix} \qquad\qquad (ii)\;\; [I]_{\mathfrak{Y}'\mathfrak{X}'} = \begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}$$

where $r = \mathrm{rk}(T) = \dim(\mathrm{range}(T))$ is the rank and $I_{r\times r}$ is the $r \times r$ identity matrix.
Hint: Finding a basis that produces (i) is fairly easy; part (ii) requires some thought about the order in which basis vectors are listed. Both matrices represent the same operator $T : V \to W$.

The matrix description of T could hardly be simpler than those in Exercise 4.2, but at the same time much information about T has been lost in allowing arbitrary unrelated bases in V and W. Most operators encode far more information than can be captured by the single number $\mathrm{rk}(T)$.

In addition to our description of a linear operator $T : V \to W$ as a matrix, we can also describe vectors $v \in V$, $w \in W$ as column matrices once bases $\mathfrak{X} = \{e_i\}$ and $\mathfrak{Y} = \{f_j\}$ are specified. The correspondence $\phi_{\mathfrak{X}} : V \to \mathbb{K}^m$ is a linear bijection (an isomorphism of vector spaces) defined by letting

$$\phi_{\mathfrak{X}}(v) = [v]_{\mathfrak{X}} = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} \qquad\qquad \text{if } v = \textstyle\sum_{i=1}^m v_i e_i \text{ (unique expansion)}$$

Similarly $\phi_{\mathfrak{Y}} : W \to \mathbb{K}^n$ is given by

$$\phi_{\mathfrak{Y}}(w) = [w]_{\mathfrak{Y}} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \qquad\qquad \text{if } w = \textstyle\sum_{j=1}^n w_j f_j \text{ (unique expansion)}$$

These coordinate descriptions of linear operators and vectors are closely related.

4.3. Proposition. *If $T : V \to W$ is a linear operator and $\mathfrak{X}$, $\mathfrak{Y}$ are bases in V, W then for all $v \in V$:*

$$\phi_{\mathfrak{Y}}(Tv) = [T]_{\mathfrak{Y}\mathfrak{X}} \cdot \phi_{\mathfrak{X}}(v)$$

or equivalently,

$$[T(v)]_{\mathfrak{Y}} = [T]_{\mathfrak{Y}\mathfrak{X}} \cdot [v]_{\mathfrak{X}} \qquad \big(\textit{an } (n\times m)\cdot(m\times 1) \textit{ matrix product}\big)$$

Thus the i^{th} component $(Tv)_i$ of $\phi_{\mathfrak{Y}}(Tv)$ is given by the familiar formula

$$(Tv)_i = \sum_{k=1}^m t_{ik} v_k \qquad \textit{for } 1 \le i \le n$$

if $v = \sum_{k=1}^m v_k e_k$ and $[T]_{\mathfrak{Y}\mathfrak{X}} = [t_{ij}]$.
Proof: If $T(e_i) = \sum_{j=1}^n t_{ji} f_j$ and $v = \sum_{k=1}^m v_k e_k$, then

$$T(v) \;=\; T\Big(\sum_{k=1}^m v_k e_k\Big) \;=\; \sum_{k=1}^m v_k T(e_k) \;=\; \sum_{k=1}^m v_k \Big(\sum_{j=1}^n t_{jk} f_j\Big) \;=\; \sum_j \Big(\sum_k t_{jk} v_k\Big) f_j$$

So, the i^{th} component $(T(v))_i$ of $\phi_{\mathfrak{Y}}\big(T(v)\big)$ is $\sum_{k=1}^m t_{ik} v_k$, as claimed. $\square$

The natural linear maps $\phi_{\mathfrak{X}} : V \to \mathbb{K}^m$, $m = \dim(V)$, and $\phi_{\mathfrak{Y}} : W \to \mathbb{K}^n$, $n = \dim(W)$, are bijective isomorphisms. Therefore a unique linear map $\tilde{T} = \phi_{\mathfrak{Y}} \circ T \circ \phi_{\mathfrak{X}}^{-1}$ is induced from $\mathbb{K}^m \rightarrow \mathbb{K}^n$ that makes the following diagram commute.

$$(17) \qquad \begin{array}{ccc} V & \xrightarrow{T} & W \\ \phi_{\mathfrak{X}} \downarrow & & \downarrow \phi_{\mathfrak{Y}} \\ \mathbb{K}^m & \xrightarrow{\tilde{T}} & \mathbb{K}^n \end{array}$$

Figure 2.3. The diagram commutes, with $\phi_{\mathfrak{Y}} \circ T = \tilde{T} \circ \phi_{\mathfrak{X}}$.

The map $\tilde{T}$ we get when $T : V \rightarrow W$ is transferred over to a map between coordinate spaces is precisely the multiplication operator $L_A : \mathbb{K}^m \rightarrow \mathbb{K}^n$, where

$$A = [T]_{\mathfrak{Y}\mathfrak{X}}$$

is the coordinate matrix that describes T as in (15) and (16), see Exercise 4.13 below.

Once bases $\mathfrak{X}, \mathfrak{Y}$ are specified there is also a natural linear isomorphism between the space of linear operators $\mathrm{Hom}_{\mathbb{K}}(V, W)$ and the space of matrices $\mathrm{M}(n \times m, \mathbb{K})$

4.4. Lemma. *If $\mathfrak{X}$, $\mathfrak{Y}$ are bases for finite dimensional vector spaces V, W the map ϕ from $\mathrm{Hom}_{\mathbb{K}}(V, W) \rightarrow \mathrm{M}(n \times m, \mathbb{K})$ given by*

$$\phi(T) = [T]_{\mathfrak{Y}\mathfrak{X}}$$

is a linear bijection, so these vector spaces are isomorphic, and

$$\dim_{\mathbb{K}} \Big(\mathrm{Hom}_{\mathbb{K}}(V, W)\Big) = \dim_{\mathbb{K}} \Big(\mathrm{M}(n \times m, \mathbb{K})\Big) = m \cdot n$$

Proof: Linearity of ϕ follows because if $\mathfrak{X} = \{e_i\}$ and $\mathfrak{Y} = \{f_j\}$ we have

$$(\lambda \cdot T)(e_i) = \lambda \cdot (T(e_i)) = \lambda \cdot \big(\sum_{j=1}^{n} t_{ji} f_j\big) = \sum_{j=1}^{m} (\lambda t_{ji}) f_j$$

for $i \leq i \leq m$, which means that $[\lambda T]_{ij} = \lambda \cdot [T]_{ij}$. Similarly, if we write $[T_k]_{\mathfrak{Y}\mathfrak{X}} = \big[t_{ij}^{(k)}\big]$ for $k = 1, 2$ we get

$$(T_1 + T_2)(e_i) = T_1(e_i) + T_2(e_i) = \sum_j t_{ji}^{(1)} f_j + \sum_j t_{ji}^{(2)} f_j = \sum_j \,\big(t_{ji}^{(1)} + t_{ji}^{(2)}\big)\, f_j$$

So $[T_1 + T_2]_{ij} = [T_1]_{ij} + [T_2]_{ij}$, proving linearity of ϕ as a map from operators to matrices.

One-to-One: To see ϕ is one-to-one it now suffices to show $\ker(\phi) = (0)$ – i.e. if $\phi(T) = [T]_{\mathfrak{Y}\mathfrak{X}} = [0]$ then T is the zero operator on V. This is clear: If $t_{ji} = 0$ for all i, j then $T(e_i) = \sum_{j=1}^{m} t_{ji} e_j = 0$ for all i, and $T(v) = 0$ for all v because $\{e_i\}$ is a basis.

Surjective: To prove ϕ surjective: given an $n \times m$ matrix $A = [a_{ij}]$ we must produce a linear operator $T : V \rightarrow W$ and bases $\mathfrak{X}, \mathfrak{Y}$ such that $[T]_{\mathfrak{Y}\mathfrak{X}} = [a_{ij}]$. This can done by working by the definition of $[T]_{\mathfrak{Y}\mathfrak{X}}$ backward: we saw earlier that there is a unique linear operator $T : V \rightarrow W$ such that $T(e_i) = \sum_{j=1}^{n} a_{ji} f_j$, because $\{e_i\} = \mathfrak{X}$ is a basis in V. Then, by definition of $[T]_{\mathfrak{Y}\mathfrak{X}}$ as in (15) - (16) we have $t_{ij} = a_{ij}$. $\square$

When $V = W$ composition of operators $S \circ T$ makes sense and the space of linear operators $\mathrm{Hom}_{\mathbb{K}}(V, V)$ becomes a (noncommutative) *associative algebra*, with the identity operator $I = \mathrm{id}_V$ as the multiplicative identity element. The set of matrices $\mathrm{M}(n, \mathbb{K})$ is also an associative algebra, under matrix multiplication; its identity element is the $n \times n$ diagonal identity matrix $I_{n \times n} = \mathrm{diag}(1, 1, \ldots, 1)$. These systems are "isomorphic" as associative algebras, as well as vector spaces, because the bijection $\phi : \mathrm{Hom}(V, V) \rightarrow \mathrm{M}(n, \mathbb{K})$ intertertwines the multiplication operations $(\circ)$ and $(\cdot)$.

4.5. Proposition. *The bijective linear map $\phi : \mathrm{Hom}_{\mathbb{K}}(V, V) \rightarrow \mathrm{M}(n, \mathbb{K})$ intertwines the product operations in these algebras:*

$$\phi(S \circ T) = \phi(S) \cdot \phi(T) \qquad \text{for all } S, T \in \mathrm{Hom}(V, V) \tag{18}$$

Given the correspondence between operators and their matrix representations, this is equivalent to saying that

$$[S \circ T]_{\mathfrak{X}\mathfrak{X}} = [S]_{\mathfrak{X}\mathfrak{X}} \cdot [T]_{\mathfrak{X}\mathfrak{X}}$$

for every basis $\mathfrak{X}$ in V, where we take matrix product on the right.

This is a special case of a much more general result.

4.6. Proposition. *Let $U \xrightarrow{T} V \xrightarrow{S} W$ be linear maps and let $\mathfrak{X} = \{u_i\}$, $\mathfrak{Y} = \{v_i\}$, $\mathfrak{Z} = \{w_i\}$ be bases in U, V, W. Then the correspondence between operators and their matrix realizations is "covariant" in the sense that*

$$[S \circ T]_{\mathfrak{Z}\mathfrak{X}} \;=\; [S]_{\mathfrak{Z}\mathfrak{Y}} \cdot [T]_{\mathfrak{Y}\mathfrak{X}}$$

(*a matrix product of compatible non-square matrices*).

Proof: We have $S \circ T(u_i) = \sum_k (S \circ T)_{ki} w_k$ by definition, and also

$$\begin{aligned} S \circ T(u_i) &= S\big(T(u_i)\big) \;=\; S\Big(\sum_j t_{ji} v_j\Big) \;=\; \sum_j t_{ji} S(v_j) \\ &= \sum_j \Big(\sum_k s_{kl} w_k\Big)\, t_{ji} \;=\; \sum_k \Big(\sum_j s_{kj} t_{ji}\Big)\, w_k \\ &= \sum_k \big([S][T]\big)_{ki} w_k \qquad \text{(definition of matrix product)} \end{aligned}$$

Thus $[S \circ T]_{ki} = ([S][T])_{ki}$, for all i, k. $\square$

4.7. Exercise. The $n \times m$ matrices E_{ij} with a "1" in the (i, j) spot and zeros elsewhere, are a basis for matrix space $\mathrm{M}(n \times m, \mathbb{K})$ since $[a_{ij}] = \sum_{i,j} a_{ij} E_{ij}$. When $m = n$ the matrices E_{ij} have useful algebraic properties. Prove that:

1. These matrices satsify the identities

$$E_{ij} E_{k\ell} = \delta_{jk} \cdot E_{i\ell}$$

   where δ_{rs} is the **Kronecker delta symbol**, equal to 1 if $r = s$ and zero otherwise.

2. The "diagonal" elements E_{ii} are projections, with $E_{ii}^2 = E_{ii}$.

3. $E_{11} + \ldots + E_{nn} = I_{n \times n}$ (the identity matrix).

If $T : V \rightarrow W$ is an invertible (bijective) linear operator between finite dimensional spaces, then $\dim(V) = \dim(W) = n$ and the inverse map $T^{-1} : W \rightarrow V$ is also linear (recall Exercise II.2.3). From the definition of the inverse map T^{-1} we have

$$T^{-1} \circ T = \mathrm{id}_V \qquad \text{and} \qquad T \circ T^{-1} = \mathrm{id}_W \ ,$$

so each operator undoes the action of the other. For any bases $\mathfrak{X}, \mathfrak{Y}$ in V, W the corresponding matrix realizations of T and T^{-1} are inverses of each other too. To see why, first recall

$$[\mathrm{id}_V]_{\mathfrak{X}\mathfrak{X}} = I_{n \times n} \qquad \text{and} \qquad [\mathrm{id}_W]_{\mathfrak{Y}\mathfrak{Y}} = I_{n \times n} \ .$$

Then by Proposition 4.6,

$$[T^{-1}]_{\mathfrak{X}\mathfrak{Y}} \cdot [T]_{\mathfrak{Y}\mathfrak{X}} = I_{n\times n} \qquad \text{and} \qquad [T]_{\mathfrak{Y}\mathfrak{X}} \cdot [T^{-1}]_{\mathfrak{X}\mathfrak{Y}} = I_{n\times n} \ ,$$

which means that $[T^{-1}]_{\mathfrak{X}\mathfrak{Y}}$ is the inverse $[T]^{-1}_{\mathfrak{Y}\mathfrak{X}}$ of the matrix of T. When $V = W$ and there is just one basis $\mathfrak{X}$ and all this reduces to the simpler statement $[T^{-1}] = [T]^{-1}$.

4.8. Exercise. Explain why *isomorphic* vector space must have the same dimension, even if one of them is infinite dimensional.

4.9. Exercise. If $T : V \to W$ is an invertible linear operator, prove that $(T^{-1})^{-1} = T$.

4.10. Exercise. If $U \xrightarrow{T} V \xrightarrow{S} W$ are invertible linear operators, explain why $S \circ T : U \to W$ is invertible, with $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$. (Note the reversal of order.)

For $A \in \mathrm{M}(n, \mathbb{K})$, we have defined the linear operator $L_A : \mathbb{K}^n \to \mathbb{K}^n$ via $L_A(\mathbf{x}) = A{\cdot}\mathbf{x}$, regarding vectors $\mathbf{x}$ as $n \times 1$ column matrices.

4.11. Exercise. Prove that the correspondence $L : \mathrm{M}(n, \mathbb{K}) \to \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^n)$ has the following algebraic properties.

1. $L_{A+B} = L_A + L_B$ and $L_{\lambda\cdot A} = \lambda{\cdot}L_A$ for all $\lambda \in \mathbb{K}$;
2. $L_{AB} = L_A \circ L_B$;
3. If $I = I_{n\times n}$ is the identity matrix, then $L_I = \mathrm{id}_{\mathbb{K}^n}$.
4. L_A is an invertible linear operator if and only if the matrix inverse A^{-1} exists in $\mathrm{M}(n, \mathbb{K})$, and then we have $(L_A)^{-1} = \big(L_{A^{-1}}\big)$.

4.12. Exercise. Explain why the correspondence $L : \mathrm{M}(n, \mathbb{K}) \to \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^n)$ is a linear bijection and an isomorphism between these associative algebras.

4.13. Exercise. If $A \in \mathrm{M}(n \times m, \mathbb{K})$ and $\mathfrak{X}$, $\mathfrak{Y}$ are the *standard bases* in coordinate spaces $\mathbb{K}^m, K^m$ prove that the matrix $B = \big[L_A\big]_{\mathfrak{Y},\mathfrak{X}}$ that describes $L_A : \mathbb{K}^m \to \mathbb{K}^n$ for this particular choice of bases is just the original matrix A
Note: Does this work for arbitrary bases in $\mathbb{K}^n$?

4.14. Exercise. Let $\mathcal{P} = \mathbb{K}[x]$ be the infinite dimensional space of polynomials over $\mathbb{K}$. Consider the linear operators

1. Derivative: $D(a_0 + a_1x + ... + a_nx^n) = a_1 + 2a_2x + ... + na_nx^{n-1}$;
2. Antiderivative: $A(a_0 + a_1x + ... + a_nx^n) = a_0x + \frac{1}{2}a_1x^2 + \ldots + \frac{1}{n+1}a_nx^{n+1}$

Show that $D \circ A = \mathrm{id}_{\mathcal{P}}$ but that $A \circ D \neq \mathrm{id}_{\mathcal{P}}$. (What is $A \circ D$?) Show that D is surjective and A is one-to-one, but $\ker(D) \neq (0)$ and the range $R(D) \neq \mathcal{P}$.

This behavior is possible only in an infinite dimensional space. We have already observed (recall Corollary II.1.6) that if finite dimensional spaces V, W *have the same dimension*, the following statements are equivalent.

T is one-to-one $\qquad$ T is surjective $\qquad$ T is bijective

4.15. Exercise. If $A \in \mathrm{M}(n, \mathbb{K})$ define its **trace** to be

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} \qquad \text{(sum of the diagonal entries)}$$

Show that, for any $A, B \in \mathrm{M}(n, \mathbb{K})$,

1. $\mathrm{Tr} : \mathrm{M}(n, \mathbb{K}) \rightarrow \mathbb{K}$ is a linear map.
2. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$;
3. If $B = SAS^{-1}$ for some invertible matrix $S \in \mathrm{M}(n, \mathbb{K})$ then $\mathrm{Tr}\big(SAS^{-1}\big) = \mathrm{Tr}(A)$.

Change of Basis and Similarity Transformations. If $T : V \rightarrow W$ is a linear map between finite dimensional spaces and $\mathfrak{X}, \mathfrak{X}' \subseteq V$ and $\mathfrak{Y}, \mathfrak{Y}' \subseteq W$ are different bases, it is important to understand how the matrix models $[T]_{\mathfrak{Y}\mathfrak{X}}$ and $[T]_{\mathfrak{Y}'\mathfrak{X}'}$ are related as we seek particular bases yielding simple descriptions of T. For instance if $T : V \rightarrow V$ we may ask if T is diagonalizable over $\mathbb{K}$: Is there a basis such that

$$[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} \lambda_1 & & & 0 \\ 0 & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

(repeats allowed among the λ_i). Not all operators are so nice, and if T is not diagonalizable we will eventually work out a satisfactory but more complicated "Plan B" for dealing with such operators. All this requires a clear understanding of how matrix descriptions behave under a "change of basis."

4.16. Theorem (Change of Basis). *Let $T : V \rightarrow V$ be a linear operator on a finite dimensional space and let id_V be the identity operator on V. If $\mathfrak{X} = \{e_1, ..., e_n\}$ and $\mathfrak{Y} = \{f_1, ..., f_n\}$ are bases in V, then*

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}} \cdot [T]_{\mathfrak{X}\mathfrak{X}} \cdot [\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}} \tag{19}$$

Futhermore $[\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}$ and $[\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}}$ are inverses of each other.

Proof: Since $T = \mathrm{id}_V \circ T \circ \mathrm{id}_V : V \rightarrow V \rightarrow V$, repeated application of Proposition 4.3 yields (19), as in the following system of commuting diagrams.

$$\begin{array}{ccccccc}
\mathbb{K}^n & \xrightarrow{[\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}} & \mathbb{K}^n & \xrightarrow{L_A} & \mathbb{K}^n & \xrightarrow{[\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}}} & \mathbb{K}^n \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \\
(V, \mathfrak{Y}) & \xrightarrow{\mathrm{id}_V} & (V, \mathfrak{X}) & \xrightarrow{T} & (V, \mathfrak{X}) & \xrightarrow{\mathrm{id}_V} & (V, \mathfrak{Y})
\end{array}$$

where $A = [T]_{\mathfrak{X}\mathfrak{X}}$. Applying the same proposition to the maps $\mathrm{id}_V = \mathrm{id}_V \circ \mathrm{id}_V$ we get

$$I_{n \times n} = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{Y}} = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}} \cdot [\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}$$

which proves $[\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}$ and $[\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}}$ are mutual inverses. $\square$

To summarize: there is a unique, invertible "transition matrix" $S \in \mathrm{M}(n, \mathbb{K})$ such that

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = S \cdot [T]_{\mathfrak{X}\mathfrak{X}} \cdot S^{-1}, \qquad \text{where} \qquad S = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}} \text{ and } S^{-1} = [\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}} \tag{20}$$

If we have explicit vector equations expressing the $\mathfrak{Y}$-basis vectors in terms of the $\mathfrak{X}$-basis vectors, the matrix $S^{-1} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}}$ can be written down immediately; then we can compute $S = \big(S^{-1}\big)^{-1}$ from it.

4.17. Definition (Similarity Transformations). *Two matrices A, B in $\mathrm{M}(n, \mathbb{K})$ are* **similar** *if there is an invertible matrix $S \in \mathrm{GL}(n, \mathbb{K})$ such that $B = SAS^{-1}$. The mapping $\sigma_S : \mathrm{M}(n, \mathbb{K}) \rightarrow \mathrm{M}(n, \mathbb{K})$ given by $\sigma_S(A) = SAS^{-1}$ is referred to as a* **similarity transformation** *of A. It is also referred to by algebraists as* **"conjugation"** *of arbitrary matrices A by an invertible matrix S.*

Each individual conjugation operator $\sigma_S(A) = SAS^{-1}$ is an *automorphism* of the associative matrix algebra – it is a bijection that respects all algebraic operations in $\mathrm{M}(n,\mathbb{K})$:

$$\begin{aligned}
\sigma_S(A\cdot B) &= \sigma_S(A)\cdot\sigma_S(B)\\
\sigma_S(A+B) &= \sigma_S(A)+\sigma_S(B)\\
\sigma_S(\lambda\cdot A) &= \lambda\cdot\sigma_S(A) \qquad \text{for } \lambda\in\mathbb{K}\\
\sigma_S\big(I_{n\times n}\big) &= I_{n\times n}
\end{aligned}$$

for all matrices A, B and all "conjugators" $S \in \mathrm{GL}(n,\mathbb{K})$. But there is even more to be said: the correspondence $\psi : S \to \sigma_S$ has important algebraic properties of its own,

$$\begin{aligned}
\sigma_{S_1S_2} &= \sigma_{S_1}\circ\sigma_{S_2} \qquad \text{for all invertible matrices } S_1, S_2\\
\sigma_{I_{n\times n}} &= \big(\text{the identity operator } \mathrm{id}_{\mathrm{M}} \text{ on matrix space } \mathrm{M}=\mathrm{M}(n,\mathbb{K})\,\big)
\end{aligned}$$

from which we automatically conclude that

The operator $\sigma_{S^{-1}}$ is the inverse $\big(\sigma_S\big)^{-1}$ of conjugation by S.

Thus the conjugation operators $\{\sigma_S : S \in \mathrm{GL}\}$ form a *group* of automorphisms acting on the algebra of $n\times n$ matrices.

When a linear operator $T : V \to V$ is described with respect to different bases in V, the resulting matrices must be similar as in (20). The converse is also true: if $A = [T]_{\mathfrak{X}\mathfrak{X}}$ and $B = SAS^{-1}$ for some invertible matrix S, there is a basis $\mathfrak{Y}$ such that $B = [T]_{\mathfrak{Y}\mathfrak{Y}}$. Thus, the different matrix models of T corresponding to bases $\mathfrak{Y}$ other than $\mathfrak{X}$ are precisely the similarity transforms $\{S\,[T]_{\mathfrak{X}\mathfrak{X}}S^{-1} : S \text{ is invertible}\}$.

4.18. Lemma. *If $T : V \to V$ is a linear operator on a finite dimensional vector space V and if $A = [T]_{\mathfrak{X}\mathfrak{X}}$ then a $n\times n$ matrix B is equal to $[T]_{\mathfrak{Y}\mathfrak{Y}}$ for some basis $\mathfrak{Y}$ if and only if $B = SAS^{-1}$ for some invertible matrix S.*

Proof: ($\Rightarrow$) follows from (20). For ($\Leftarrow$): since S is invertible it has a matrix inverse S^{-1}. (Later we will discuss effective methods to compute matrix inverses such as S^{-1}.) According to Theorem 4.16, what we need is a basis $\mathfrak{Y} = \{f_1, ..., f_n\}$ such that $[\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}} = S^{-1}$; then $[\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}} = (S^{-1})^{-1} = S$ and $B = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^{-1}$. If we write $S^{-1} = [a_{ij}]$ the identity $S^{-1} = [\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}$ means that

$$f_i = \mathrm{id}_V(f_i) = \sum_j a_{ji}e_j \qquad \text{for } 1 \le i \le n$$

where $\mathfrak{X} = \{e_1, ..., e_n\}$. This is the desired new basis $\mathfrak{Y} = \{f_j\}$. To see it is a basis, we have $\{f_j\} \subseteq \mathbb{K}\text{-span}\{e_1, ..., e_n\}$ by definition, but $\{e_i\}$ is in $\mathbb{K}\text{-span}\{f_j\}$ because the matrix $S^{-1} = [a_{ij}]$ is invertible; in fact, $S^{-1}S = I$ implies that $\sum_i a_{ji}s_{ik} = \delta_{jk}$ (Kronecker delta). Then

$$\sum_i s_{ik}f_i = \sum_i s_{ik}\Big(\sum_j a_{ji}e_j\Big) = \sum_j \Big(\sum_i a_{ji}s_{ik}\Big)\,e_j = \sum_j \delta_{jk}e_j = e_k$$

for $1 \le k \le n$, so $\{e_k\} \subseteq \mathbb{K}\text{-span}\{f_j\}$ as claimed. Therefore $\{e_1, ..., e_n\}$ and $\{f_1, ...f_n\}$ both span V, and because $\{e_i\}$ is already a basis $\{f_j\}$ must also be a basis. $\square$

The next example shows that it can be difficult to tell by inspection whether an operator T is diagonalizable.

4.19. Example. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be the linear operator whose action on the standard basis $\mathfrak{X} = \{\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1)\}$ is

$$T(\mathbf{e}_1) = 4\mathbf{e}_1 \qquad T(\mathbf{e}_2) = -\mathbf{e}_2$$

Clearly T is diagonalized by the $\mathfrak{X}$-basis since

$$[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix}$$

Compute $[T]_{\mathfrak{Y}\mathfrak{Y}}$ for the basis

$$\mathbf{f}_1 = \frac{1}{\sqrt{2}}(\mathbf{e}_1 + \mathbf{e}_2) \qquad \mathbf{f}_2 = \frac{1}{\sqrt{2}}(\mathbf{e}_1 - \mathbf{e}_2)$$

Solution: We have $[T]_{\mathfrak{Y}\mathfrak{Y}} = S\,[T]_{\mathfrak{X}\mathfrak{X}} S^{-1}$ where $S = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}}$ and $S^{-1} = [\mathrm{id}_V]_{\mathfrak{X}\mathfrak{Y}}$. This inverse can be computed easily from our definition of the vectors $\mathbf{f}_1$, $\mathbf{f}_2$:

$$\begin{aligned} \mathbf{f}_1 &= \mathrm{id}(\mathbf{f}_1) = \frac{1}{\sqrt{2}}\mathbf{e}_1 + \frac{1}{\sqrt{2}}\mathbf{e}_2 \\ \mathbf{f}_2 &= \mathrm{id}(\mathbf{f}_2) = \frac{1}{\sqrt{2}}\mathbf{e}_1 - \frac{1}{\sqrt{2}}\mathbf{e}_2 \end{aligned} \tag{21}$$

which implies

$$S^{-1} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

The inverse of this matrix (found by standard matrix algebra methods or simply by solving (21) for $\mathbf{e}_1$, $\mathbf{e}_2$ in terms of $\mathbf{f}_1$, $\mathbf{f}_2$) is

$$[\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} = (S^{-1})^{-1} = S = -\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

(Notice that $S = S^{-1}$; this is not usually the case.) Then we get

$$\begin{aligned} [T]_{\mathfrak{Y}\mathfrak{Y}} &= [S]\,[T]_{\mathfrak{X}\mathfrak{X}} S^{-1} = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \\ &= \tfrac{1}{2} \cdot \begin{pmatrix} 4 & -1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \tfrac{1}{2} \cdot \begin{pmatrix} 3 & 5 \\ 5 & 3 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{5}{2} \\ \frac{5}{2} & \frac{3}{2} \end{pmatrix} \end{aligned}$$

Diagonalizability of T would not be at all apparent if we used the basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ to represent T. $\square$

4.20. Exercise. Compute the matrix $[T]_{\mathfrak{Y}\mathfrak{Y}}$ for the linear operator $T : \mathbb{C}^2 \to \mathbb{C}^2$ of the previous example for each of the folowing bases $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$:

1. $\begin{cases} \mathbf{f}_1 = \frac{1}{\sqrt{2}}(\mathbf{e}_1 + \mathbf{e}_2) \\ \mathbf{f}_2 = \frac{1}{\sqrt{2}}(-\mathbf{e}_1 + \mathbf{e}_2) \end{cases}$ (obtained by rotating the standard basis vectors by $\theta = +45°$)

2. $\begin{cases} \mathbf{f}_1 = \frac{\sqrt{3}}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2 \\ \mathbf{f}_2 = -\frac{1}{2}\mathbf{e}_1 + \frac{\sqrt{3}}{2}\mathbf{e}_2 \end{cases}$ (the standard basis vectors rotated by $\theta = +30°$)

3. $\begin{cases} \mathbf{f}_1 = \mathbf{e}_1 + i\mathbf{e}_2 \\ \mathbf{f}_2 = \mathbf{e}_1 - i\mathbf{e}_2 \end{cases}$ (where $i = \sqrt{-1}$ in $\mathbb{C}$ and $V = \mathbb{C}^2$)

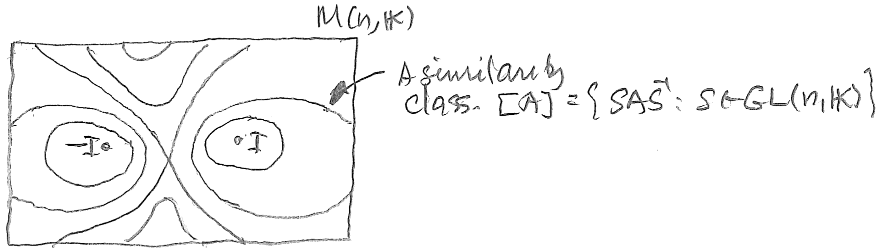

Figure 2.4. Similarity classes (= conjugacy classes) in $\mathrm{M}(n,\mathbb{K})$ partition matrix space into disjoint subsets $[A]$. Some classes are single points, for instance $[-I], [I]$, and $[0]$. Othere sre complicated hypersurfaces in $\mathbb{R}^{n^2} \cong \mathrm{M}(n,\mathbb{R})$. A similarity class could have seevral disconnected components.

4.21. Exercise. Let $\mathcal{P}_n$= polynomials of degree $\leq n$. Let $D = d/dx : \mathcal{P}_n \to \mathcal{P}_n$, the formal derivative of a polynomial. Compute $[D]_{\mathfrak{X}\mathfrak{X}}$ with respect to the basis $\mathfrak{X} = \{1, x, ..., x^n\}$. Compute $[D^2]_{\mathfrak{X}\mathfrak{X}}$ and $[D^{n+1}]_{\mathfrak{X}\mathfrak{X}}$ too.

An **RST equivalence relation** on a set X is rule declaring certain points $x, y \in X$ to be "related" (and others not). Writing $x \underset{R}{\sim} y$ when the points are related, the phrase "RST" means the relation is

1. Reflexive: $x \underset{R}{\sim} x$ for all $x \in X$.
2. Symmetric: $x \underset{R}{\sim} y \Rightarrow y \underset{R}{\sim} x$.
3. Transitive: $x \underset{R}{\sim} y$ and $y \underset{R}{\sim} z \Rightarrow x \underset{R}{\sim} z$

For each $x \in X$ we can then define its *equivalence class*, the subset

$$[x]_R = \{y \in X : y \underset{R}{\sim} x\}$$

The RST property forces distinct equivalence classes to be disjoint, so the whole space X decomposes into a the disjoint union of these classes.

One example of an RST equivalence is "*congruence mod a fixed prime* p" in the set $X = \mathbb{Z}$,

$$k \sim \ell \Leftrightarrow k \equiv \ell \pmod{p} \Leftrightarrow \ k \text{ and } \ell \text{ differ by a multiple of } p$$

It is easily verified that this is an RST relation and that the equivalence class of an integer m is its (mod p) congruence class

$$[m] = m + p\mathbb{Z} = \{k \in \mathbb{Z} : k \equiv m \text{ (mod p)}\}$$

There are only finitely many distinct classes, namely $[0], [1], \ldots, [p-1]$, which are disjoint and fill $\mathbb{Z}$. The finite field $\mathbb{Z}_p$ is precisely this set of equivalence classes equipped with suitable $\oplus$ and $\odot$ operations inherited from the system of integers $(\mathbb{Z}, +, \cdot)$.

Similarity of matrices

$$A \underset{R}{\sim} B \Leftrightarrow B = SAS^{-1} \quad \text{for some invertible matrix} \quad S \in \mathrm{GL}(n,\mathbb{K}) \tag{22}$$

is an important example of an RST relation on matrix space $X = \mathrm{M}(n,\mathbb{K})$. The RST properties are easily verified.

4.22. Exercise. Prove that similarity of matrices (22) has each of the RST properties.

The equivalence classes partition $\mathrm{M}(n, \mathbb{K})$ into disjoint **"similarity classes"** (aka **"conjugacy classes"**). All the matrices $[T]_{\mathfrak{X}\mathfrak{X}}$ associated with a linear operator $T : V \to V$ constitute a single similarity class in matrix space – they are all the possible representations of T corresponding to different choice of bases in V – and different operators correspond to disjoint similarity classes in $\mathrm{M}(n, \mathbb{K})$.

The similarity classes don't all look the same. Some are trivial, consisting of a single point: for instance if

$$A = 0 \quad \text{or} \quad A = \lambda I_{n\times n} \ \ \text{(a scalar multiple of the identity matrix)}$$

we have

$$SAS^{-1} = \lambda{\cdot}SIS^{-1} = \lambda \cdot SS^{-1} = \lambda I = A \qquad \text{for all } S \in \mathrm{GL}(n, \mathbb{K})$$

The similarity class $[A]$ consists of the single point A. In particular, $[0] = \{0\}$, $[I] = \{I\}$ and $[-I] = \{-I\}$. When $\mathbb{K} = \mathbb{R}$ and we identify $\mathrm{M}(n, \mathbb{R})$ with $\mathbb{R}^{n^2}$, other similarity classes can be large curvilinear surfaces in Euclidean space. They can be quite a mess to compute.

4.23. Exercise. If $A \in \mathrm{M}(n, \mathbb{K})$ prove that

1. A commutes with *all* $n \times n$ matrices $\Leftrightarrow A = \lambda{\cdot}I_{n\times n}$, a scalar multiple if the identity matrix for some $\lambda \in \mathbb{K}$.
2. A commutes with all matrices in all *invertible* matrices $\mathrm{GL}(n, \mathbb{K}) = \{A : \det(A) \neq 0\} \Leftrightarrow A$ commutes with *all* $n \times n$ matrices, as in (1.)

Hint: Recall the matrices E_{ij} defined in Exercise 4.7, which are a basis for matrix space. In (2.), if $i \neq j$ then $I + E_{ij}$ is invertible (verify that $(I - E_{ij})$ is the inverse), and commutes with A. Hence E_{ij} commutes with A; we leave you to figure out what to do when $i = j$. If A commutes with all basis vectors E_{ij} it obviously commutes with all $n \times n$ matrices, and (1.) can be applied.

This shows that a similarity class $[A]$ in $\mathrm{M}(n, \mathbb{K})$ consists of a single point $\Leftrightarrow A = \lambda I$ (a *scalar matrix*).

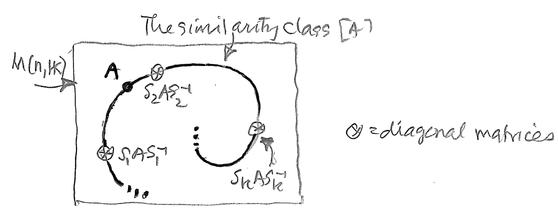

Figure 2.5. The diagonalization problem for $A \in \mathrm{M}(n, \mathbb{K})$ amounts to searching for one (or more) diagonal matrices lying in the similarity class $[A] = \{SAS^{-1} : S \in \mathrm{GL}(n, \mathbb{K})\}$.

4.24. Exercise. When we identify $\mathrm{M}(2, \mathbb{R}) \cong \mathbb{R}^4$ via the linear isomorphism

$$\mathbf{x} = \phi(A) = \big(a_{11}, a_{12}, a_{21}, a_{22}\big) \ ,$$

show that the similarity class $[A]$ of the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is the 2-dimensional surface in $\mathbb{R}^4$ whose description in parametric form, described as the range of a polynomial map $\phi : \mathbb{R}^2 \to \mathrm{M}(2, \mathbb{R})$, is

$$[A(s,t)] = \left\{ \begin{pmatrix} 1-st & s^2 \\ -t^2 & 1+st \end{pmatrix} : s,t \in \mathbb{R} \text{ and } (s,t) \neq (0,0) \right\}$$

Note: A matrix $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if the determinant $\det(S) = ad - bc$ is not 0, and then the inverse matrix is

$$S^{-1} = \frac{1}{\det(S)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \square$$

We will have a lot more to say about change of basis, similarity classes, and the diagonalization problem later on. Incidentally, not all matrices can be put into diagonal form by a similarity transformation. Our fondest hope is that in the equivalence class $[A]$ there will be at least one point SAS^{-1} that is diagonal (there may be several, as in Figure 2.5). If $A = [T]_{\mathfrak{X}\mathfrak{X}}$ for some linear operator $T : V \to V$ this is telling us which bases $\mathfrak{Y}$ make $[T]_{\mathfrak{Y}\mathfrak{Y}}$ diagonal, or whether there are any such bases at all.

4.25. Exercise. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear operator such that $T(\mathbf{e}_1) = 0$ and $T(\mathbf{e}_2) = \mathbf{e}_1$, so its matrix with respect to the standard basis $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\}$ is $[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Prove that *no* basis $\mathfrak{Y} = \{f_1, f_2\}$ can make $[T]_{\mathfrak{Y}\mathfrak{Y}}$ diagonal.

Hint: $S = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is invertible $\Leftrightarrow$ the determinant $\det(S) = a_{11}a_{22} - a_{12}a_{21}$ is nonzero. We will eventually develop systematic methods to answer questions of this sort. For the moment, you will have to do it "bare-hands."