# LINEAR TRANSFORMATIONS

TIM SMITS

A general philosophy is that to study algebraic structures, one needs to not just study the objects but structure preserving maps between these objects as well. There is no simple explanation for the latter, but historically it has been very productive. When studying how to solve systems of linear equations in $\mathbb{R}^n$, one is naturally led to matrix equations of the form $Ax = b$. A matrix $A$ defines a function $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$. This function $T$ respects the structure of Euclidean space, in the sense that $T(x + y) = T(x) + T(y)$ and $T(cx) = cT(x)$ for all $x, y \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Since vector spaces are nothing more than abstracted versions of Euclidean space, we should look at abstract analogues of matrices, i.e. functions that preserve the vector space structure.

Unless otherwise stated, through the handout $V$ is a finite dimensional vector space of dimension $n$ over a field $F$. The letter $T$ will always denote a linear transformation.

## BASIC DEFINITIONS AND EXAMPLES

**Definition 0.1.** A **linear transformation** $T : V \to W$ between vector spaces $V$ and $W$ over a field $F$ is a function satisfying $T(x + y) = T(x) + T(y)$ and $T(cx) = cT(x)$ for all $x, y \in V$ and $c \in F$. If $V = W$, we sometimes call $T$ a **linear operator** on $V$.

Note that necessarily a linear transformation satisfies $T(0) = 0$. We also see by induction that for any finite collection of vectors $v_1, \ldots, v_n$ and scalars $c_1, \ldots, c_n \in F$ we have $T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n)$.

**Definition 0.2.** The **kernel** $\ker(T)$ is defined by $\ker(T) = \{x \in V : T(x) = 0\}$. The **image** $\mathrm{Im}(T)$ is defined by $\mathrm{Im}(T) = \{T(x) : x \in V\}$.

The image and kernel of $T$ are two important subspace of $V$ and $W$ respectively, and we can translate set theoretic statements about injectivity and surjectivity into the language of linear algebra.

**Proposition 1.** *Let $T : V \to W$ be linear. Then $\ker(T)$ is a subspace of $V$ and $\mathrm{Im}(T)$ is a subspace of $W$.*

*Proof.* Since $T$ is linear, we have $T(0) = T(0 + 0) = T(0) + T(0)$, so $0 = T(0)$ gives $0 \in \ker(T)$. If $x, y \in \ker(T)$ then $T(x + y) = T(x) + T(y) = 0$ by linearity. Similarly, if $c \in F$, $T(cx) = cT(x) = 0$ so $cx \in \ker(T)$ giving $\ker(T)$ is a subspace of $V$. Since $T(0) = 0$, this says $0 \in \mathrm{Im}(W)$. If $x, y \in \mathrm{Im}(W)$ then there are $u, v \in V$ such that $x = T(u)$ and $y = T(v)$. Then $x + y = T(u) + T(v) = T(u + v)$ so $x + y \in \mathrm{Im}(T)$. Finally, if $x = T(u)$ then $cx = cT(u) = T(cu)$ so $cx \in \mathrm{Im}(T)$ which says $\mathrm{Im}(T)$ is a subspace of $W$. □

**Proposition 2.** *Let $T : V \to W$ be linear.*

(a) *$T$ is injective if and only if $\ker(T) = \{0\}$.*
(b) *$T$ is surjective if and only if $\mathrm{Im}(T) = W$.*

*Proof.*

(a) Suppose that $T$ is injective. If $x \in \ker(T)$, then $T(x) = T(0)$ so injectivity says $x = 0$ giving $\ker(T) = \{0\}$. If $\ker(T) = \{0\}$, if $T(x) = T(y)$ then $T(x - y) = 0$ says $x - y \in \ker(T)$ so $x - y = 0$, i.e. $x = y$ so $T$ is injective.

(b) If $T$ is surjective, then for every $y \in W$ there is $x \in V$ such that $T(x) = w$, which is precisely the same as saying $W = \mathrm{Im}(T)$. On the other hand, if $W = \mathrm{Im}(T)$ then for all $w \in W$ there is $x \in V$ with $w = T(x)$, so $T$ is surjective.

□

**Example 0.3.** For any vector space $V$, the **identity transformation** $\mathrm{id}_V : V \to V$ given by $\mathrm{id}_V(x) = x$ is linear.

**Example 0.4.** For any field $F$ and $a \in F$, the map $T : F \to F$ given by $T(x) = ax$ is a linear transformation by the field axioms.

**Example 0.5.** The map $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(x, y) = (x + y, y, x - y)$ is a linear transformation.

**Example 0.6.** For any matrix $A \in M_{m\times n}(F)$, the map $T : F^n \to F^m$ given by $T(x) = Ax$ is a linear transformation, since $A(x + y) = Ax + Ay$ and $A(cx) = c(Ax)$ by how matrices work.

**Example 0.7.** In $P(\mathbb{R})$, the maps $D(p)(x) = p'(x)$ and $I(p)(x) = \int_0^x p(t)\,dt$ are linear operators on $P(\mathbb{R})$ by calculus. $D$ is not injective, because any constant polynomial has derivative 0, but $D$ is surjective since $D(\int_0^x p(t)\,dt) = p(x)$ by the fundamental theorem of calculus. The operator $I$ is injective but not surjective because nothing maps to the polynomial $p(x) = 1$.

**Example 0.8.** The map $D : C^\infty([0, 1]) \to C^\infty([0, 1])$ given by $D(f)(x) = f(x) - f'(x)$ is a linear transformation. Saying $f \in \ker(D)$ is the same as saying $f'(x) = f(x)$, so $\ker(D)$ is precisely the set of functions that satisfy the differential equation $f = f'$. From calculus, we know the only such functions are of the form $ce^x$ for $c \in \mathbb{R}$, so $\ker(D) = \mathrm{Span}(\{e^x\})$ is a 1-dimensional subspace of $C^\infty([0, 1])$.

**Example 0.9.** The map $T : M_n(F) \to M_n(F)$ given by $T(A) = A - A^t$ is linear. $\ker(T)$ is the set of matrices with $A = A^t$, i.e. $\ker(T) = \mathrm{Sym}_n(F)$. Any matrix in $\mathrm{Im}(T)$ is of the form $A - A^t$ for some $A$, which is skew-symmetric, so $\mathrm{Im}(T) \subset \mathrm{Skew}_n(F)$. If $F$ does not have characteristic 2, for any skew-symmetric matrix $B$, we have $T(B) = B - B^t = 2B$, so $T(\frac{1}{2}B) = B$ says $\mathrm{Im}(T) = \mathrm{Skew}_n(F)$.

**Example 0.10.** Let $F^\infty$ be the sequence space of elements of $F$. That is, $F^\infty = \{(a_1, a_2, \ldots) : a_i \in F\}$. Define maps $R : F^\infty \to F^\infty$ by $R((a_1, a_2, \ldots)) = (a_2, a_3, \ldots)$ and $L((a_1, a_2, \ldots)) = (0, a_1, a_2, \ldots)$, the right and left shift operators respectively. Then both $R$ and $L$ are linear operators on $F^\infty$.

**Example 0.11.** Let $W \subset V$ be a subspace. The map $\pi : V \to V/W$ given by $T(v) = v + W$ is a linear transformation, called the **quotient map**.

**Example 0.12.** Suppose $V = W \oplus U$ for some subspaces $W, U$ of $V$. The **projection** $\pi_W$ of $V$ onto $W$ along $U$ is defined by $\pi_W(x) = w$ where $x = w + u$ for unique $w \in W$ and $u \in U$. Then $\pi_W$ is linear, and $\ker(\pi_W) = U$ and $\mathrm{Im}(\pi_W) = W$. If we assume $V$ is finite dimensional, for any subspace $W$ there is $U$ such that $V = W \oplus U$. This then says that any subspace $W$ is the kernel of some linear transformation, namely $\pi_U$ where $U$ is the complement of $W$ in $V$. Similarly, $W$ appears as the image of $\pi_W$.

A natural question is given a linear operator $T : V \to V$ and a subspace $W$ of $V$, when does $T$ restrict to a linear operator on $W$? Necessarily, if $T$ restricts to an operator on $W$ we must have $T(W) \subset W$, and actually this is sufficient: if $T(W) \subset W$, then for $x, y \in W$ we have $T(x+y) = T(x) + T(y)$ since $x, y \in V$ and $T(cx) = cT(x)$ for $c \in F$. We give such subspaces a name:

**Definition 0.13.** Given a linear operator $T : V \to V$, a subspace $W \subset V$ is called **T-invariant** if $T(W) \subset W$. The restriction of $T$ to $W$, denoted by $T|_W$, is the linear transformation $T|_W(x) = T(x)$ for all $x \in W$.

**Example 0.14.** Let $V = W \oplus U$, and consider $\pi_W$. Then $W$ is $\pi_W$-invariant, and $\pi_W|_W$ is the identity map on $W$.

**Example 0.15.** The map $T : M_n(F) \to M_n(F)$ given by $T(A) = A - A^t$ is $\mathrm{Sym}_n(F)$-invariant. The restriction $T|_{\mathrm{Sym}_n(F)}$ is simply the 0 map.

## Dimension counting

The image and kernel of a linear transformation $T$ are extremely important because they give rise to a powerful dimension counting result. We'll illustrate how we can use such counting arguments to get useful results.

**Definition 0.16.** The **rank** of a linear transformation $T : V \to W$, denoted by $\mathrm{rank}(T)$ is defined as $\mathrm{rank}(T) = \dim(\mathrm{Im}(T))$.

**Proposition 3.** *If $\{v_1, \dots, v_n\}$ is a basis of $V$, then $\{T(v_1), \dots, T(v_n)\}$ is a spanning set of $\mathrm{Im}(T)$. If $T$ is injective then $\{T(v_1), \dots, T(v_n)\}$ is a basis of $\mathrm{Im}(T)$. In particular, if $T$ is injective then $\mathrm{rank}(T) = n$.*

*Proof.* Let $y \in \mathrm{Im}(T)$. Then $y = T(x)$ for some $x \in V$, and we may write $x = c_1v_1 + \ldots + c_nv_n$ for some $c_i \in F$. Then $y = T(x) = T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n)$, so $y \in \mathrm{Span}(\{T(v_1), \dots, T(v_n)\}$, which says this is a spanning set of $\mathrm{Im}(T)$. If further we assume that $T$ is injective, if $c_1T(v_1) + \ldots + c_nT(v_n) = 0$, then $T(c_1v_1 + \ldots + c_nv_n) = 0$, so $c_1v_1 + \ldots + c_nv_n \in \ker(T)$. Since $T$ is injective, this says $c_1v_1 + \ldots + c_nv_n = 0$, and since the vectors $v_i$ are linearly independent this says $c_i = 0$, i.e. that $\{T(v_1), \dots, T(v_n)\}$ is linearly independent and therefore a basis of $\mathrm{Im}(T)$, so that $\mathrm{rank}(T) = n$. □

**Theorem 0.17.** *(Rank-Nullity) Let $W$ be a vector space. If $T : V \to W$ is linear, then $\mathrm{rank}(T) + \dim(\ker(T)) = n$.*

*Proof.* Pick a basis $\{v_1, \dots, v_k\}$ of $\ker(T)$ and extend this to a basis $\{v_1, \dots, v_k, w_1, \dots, w_\ell\}$ of $V$. The above shows that $\mathrm{Im}(T) = \mathrm{Span}(\{T(w_1), \dots, T(w_\ell)\})$, so $\{T(w_1), \dots, T(w_\ell)\}$ is a spanning set and therefore it's sufficient to prove it is a basis of $\mathrm{Im}(T)$. Suppose $c_1T(w_1) + \ldots + c_\ell T(w_\ell) = 0$. Then $T(c_1w_1 + \ldots + c_\ell w_\ell) = 0$, so $c_1w_1 + \ldots + c_\ell w_\ell \in \ker(T)$. We may then write $c_1w_1 + \ldots + c_\ell w_\ell = a_1v_1 + \ldots + a_kv_k$ for some $a_i \in F$, so $c_1w_1 + \ldots + c_\ell w_\ell - a_1v_1 - \ldots - a_kv_k = 0$. Since the vectors $w_i, v_j$ are a basis of $V$, this says all $c_i = 0$ and all $a_i = 0$, so that $\{T(w_1), \dots, T(w_\ell)\}$ is a basis as desired. □

Sometimes $\dim(\ker(T))$ is referred to as the nullity of $T$, hence the name of the theorem, but this terminology is not commonly used outside of linear algebra textbooks. As an immediate corollary of the rank-nullity theorem, we getting the following analogous result for functions between finite sets of the same size:

**Corollary 0.18.** *Let $V$ and $W$ be vector spaces of the same dimension. Then $T : V \to W$ is injective $\iff$ $T$ is surjective $\iff$ $T$ is bijective.*

*Proof.* $T$ is injective if and only if $\ker(T) = \{0\}$, so by rank-nullity this says $n = \operatorname{rank}(T) + 0$, i.e. $\operatorname{Im}(T) = W$ so $T$ is surjective. Similarly if $T$ is surjective, $\operatorname{rank}(T) = n$ so rank-nullity says $n = n + \dim(\ker(T))$ so $\dim(\ker(T)) = 0$ gives $\ker(T) = \{0\}$ and therefore $T$ is injective. $\square$

We give some examples to illustrate how the rank-nullity theorem is used to compute images and kernels of linear transformations.

**Example 0.19.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (x+y+2z, 2x+2y+4z, 2x+3y+5z)$. Then $T$ is a linear transformation, and $\operatorname{Im}(T) = \operatorname{Span}(\{(1,2,2),(1,2,3),(2,4,5)\}) = \operatorname{Span}(\{(1,2,2),(1,2,3)\})$. The latter set is a basis for $\operatorname{Im}(T)$, so that $\operatorname{rank}(T) = 2$, i.e. $\operatorname{Im}(T)$ is a plane in $\mathbb{R}^3$. By rank-nullity the kernel of $T$ is 1-dimensional, so it must be a line. Which line is it? Representing $T$ as the matrix $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 2 & 3 & 5 \end{pmatrix}$, one sees that any vector orthogonal to the rows of $A$ is contained in the kernel. Taking the cross product of the first and third rows shows $(-1, 1, 1) \in \ker(T)$ so that $\ker(T) = \operatorname{Span}(\{(-1, 1, 1)\})$.

**Example 0.20.** Let $T : M_n(F) \to F$ be the trace map $T(A) = \operatorname{tr}(A)$. Clearly $T$ is surjective, so by rank-nullity we have $\dim(\ker(T)) = n^2 - 1$. For $A \in \ker(T)$, we have $a_{11} + \ldots + a_{nn} = 0$, which says $a_{11} = -a_{22} - \ldots - a_{nn}$. Since the condition on trace has nothing to do with entries off the diagonals, we see that the matrices $E_{ij}$ with $i \neq j$ along the matrices $-E_{11} + E_{ii}$ with $2 \leq i \leq n$ form a spanning set for $\ker(T)$, and therefore a basis because there are $n^2 - 1$ of them.

**Example 0.21.** The map $T : P(\mathbb{R}) \to P(\mathbb{R})$ defined by $T(p) = 5p'' + 3p'$ is surjective. Let $q$ be a polynomial of degree $n$. Restricting $T$ to $P_{n+1}(\mathbb{R})$ defines a map $T' : P_{n+1}(\mathbb{R}) \to P_n(\mathbb{R})$. with $T'(p) = T(p)$. By rank-nullity, $\operatorname{rank}(T') + \dim(\ker(T')) = n + 2$. If $p \in \ker(T')$, then $5p'' + 3p' = 0$ says $5p'' = -3p'$. Since $\deg(p'') = \deg(p') - 1$, this is impossible unless both $p'$ and $p''$ are 0, i.e. $p$ is a constant. This says $\ker(T') = \operatorname{Span}(\{1\})$, so $\ker(T')$ is 1-dimensional, and $\operatorname{rank}(T') = n + 1$ says $T'$ is surjective.

**Example 0.22.** Let $F$ not be of characteristic 2, and $T : M_n(F) \to M_n(F)$ be given by $T(A) = A - A^t$. The rank-nullity theorem says that $\dim(\operatorname{Sym}_n(F)) + \dim(\operatorname{Skew}_n(F)) = n^2$. One can check (by explicitly finding a basis) that $\dim(\operatorname{Skew}_n(F)) = \frac{n(n-1)}{2}$, so $\dim(\operatorname{Sym}_n(F)) = n^2 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$.

**Example 0.23.** Let $\dim(V) = n$ and $\dim(W) = m$ with $n < m$. Then there is no surjective linear transformation $T$ from $V$ to $W$: by rank-nullity, $\operatorname{rank}(T) + \dim(\ker(T)) = n$, so $\operatorname{rank}(T) = n - \dim(\ker(T)) \leq n < m$ says $\operatorname{Im}(T) \neq W$. Similarly, if $n > m$ there is no injective linear transformation from $V$ to $W$. This says if $n < m$ then an $m$-dimensional vector space is "larger" than an $n$-dimensional vector space, which hopefully matches with your intuition.

We end with some useful dimension formulas:

**Proposition 4.** *Let $W, U \subset V$ be subspaces. Then* $\dim(W \oplus U) = \dim(W) + \dim(U)$ *and* $\dim(V/W) = \dim(V) - \dim(W)$.

*Proof.* Define $T : W \oplus U \to W$ by $x \to x_W$, where the element $x \in W \oplus U$ is written uniquely as $x_W + x_U$ for some $x_W \in W$ and $x_U \in U$. This map is easily see to be linear, and is surjective since $T(w + 0) = w$ for any $w \in W$. We also see that $\ker(T) = U$, so by rank-nullity, we have $\dim(W \oplus U) = \dim(U) + \dim(W)$ as desired.

For the other statement, define $T : V \to V/W$ by $T(v) = v + W$. Then $T$ is linear, and $T$ is clearly surjective by the way it's defined. We also see that $\ker(T) = W$, so that rank-nullity says $\dim(V) = \dim(W) + \dim(V/W)$. ☐

## The vector space $\mathrm{Hom}_F(V, W)$

**Definition 0.24.** A linear transformation $T : V \to W$ is called an **isomorphism** if $T$ is bijective. $V$ and $W$ are called **isomorphic** if there is an isomorphism between them, and we write $V \cong W$.

**Definition 0.25.** Let $V$ and $W$ be $F$-vector spaces. Then we define $\mathrm{Hom}_F(V, W) = \{T : V \to W : T \text{ is linear}\}$, the set of linear transformations from $V$ to $W$. If $V = W$, we instead write $\mathrm{End}_F(V)$.

**Proposition 5.** *$Hom_F(V, W)$ is a subspace of $\mathcal{F}(V, W)$.*

*Proof.* If $T, U \in \mathrm{Hom}_F(V, W)$ recall that by definition, we have $(T + U)(v) = T(v) + U(v)$ and $(cT)(v) = cT(v)$. To check that $\mathrm{Hom}_F(V, W)$ is a subspace, we need to check that it is non-empty, and that for $T, U$ linear transformations and $c \in F$ that $T + U$ and $cT$ are linear transformations. Note that the 0 function $T(x) = 0$ for all $x \in V$ is certainly linear. If $x, y \in V$ then $(T + U)(x + y) = T(x + y) + U(x + y) = T(x) + U(x) + T(y) + U(y) = (T + U)(x) + (T + U)(y)$. Next, for $c \in F$ we have $(T + U)(cx) = T(cx) + U(cx) = cT(x) + cU(x) = c(T(x) + U(x)) = c(T + U)(x)$, so $T + U$ is a linear transformation which says $T+U \in \mathrm{Hom}_F(V, W)$. For $x, y \in V$, $c, k \in F$, we have $(cT)(x+y) = (cT)(x)+(cT)(y) = cT(x) + cT(y) = (cT)(x) + (cT)(y)$, and $(cT)(kx) = cT(kx) = k(cT(x)) = k(cT)(x)$. This says $cT$ is a linear transformation, so $cT \in \mathrm{Hom}_F(V, W)$ so that $\mathrm{Hom}_F(V, W)$ is a subspace of $\mathcal{F}(V, W)$. ☐

In most linear algebra books the space $\mathrm{Hom}_F(V, W)$ is denoted as $\mathcal{L}(V, W)$ and $\mathrm{End}_F(V)$ as $\mathcal{L}(V)$, but the above notation is more common elsewhere in mathematics. One of the reasons why finite dimensional vector spaces are so easy to study is that linear transformations between $V$ and $W$ are the same things as functions defined on a basis of $V$. This reduces much of the study of linear algebra to studying functions defined on a finite set. This is stated precisely in the following form.

**Theorem 0.26.** *Let $V$ be finite dimensional, and let $B$ be a basis of $V$. Then $Hom_F(V, W) \cong \mathcal{F}(B, W)$. In other words, every linear transformation is determined uniquely by what it does on a basis of $V$.*

*Proof.* Suppose that $T : V \to W$ is a linear transformation, and let $B = \{v_1, \ldots, v_n\}$ be a basis of $V$. For $x \in V$, we may uniquely write $x = c_1v_1 + \ldots + c_nv_n$, so $T(x) = c_1T(v_1) + \ldots +$

$c_nT(v_n)$ by linearity. This defines a function $f : B \to W$ by $f(v_i) = T(v_i)$. Now suppose we have a function $f : B \to W$. Define $T_f : V \to W$ by $T_f(c_1v_1 + \ldots + c_nv_n) = c_1f(v_1) + \ldots + c_nf(v_n)$. We need to check that $T_f$ is a linear transformation, and that it is the only transformation that agrees with $f$ on $B$. Write $x = c_1v_1+\ldots+c_nv_n$ and $y = d_1v_1+\ldots+d_nv_n$. Then $T_f(x+y) = T_f((c_1+d_1)v_n+\ldots+(c_n+d_n)v_n) = (c_1+d_1)f(v_1)+\ldots+(c_n+v_n)f(v_n) = c_1f(v_1)+\ldots+c_nf(v_n)+d_1f(v_1)+\ldots+d_nf(v_n) = T_f(x)+T_f(y)$. For $c \in F$, we have $T_f(cx) = T_f((cc_1)v_1+\ldots+(cc_n)v_n) = (cc_1)f(v_1)+\ldots+(cc_n)f(v_n) = c(c_1f(v_1)+\ldots+c_nf(v_n)) = cT_f(x)$, which shows that $T_f$ is linear. Finally, suppose there is some other linear transformation $T' : V \to W$ such that $T'(v_i) = f(v_i)$. As mentioned above then says for any $x \in V$, $T'(x) = T'(c_1v_1+\ldots+c_nv_n) = c_1T'(v_1)+\ldots+c_nT'(v_n) = c_1f(v_1)+\ldots c_nf(v_n) = T_f(x)$, i.e. $T' = T_f$ so $T_f$ is the only linear transformation with this property.

Putting this all together, this says the map $G : \mathcal{F}(B,W) \to \mathrm{Hom}_F(V,W)$ with $G(f) = T_f$ is a bijection: it is injective because if $G(f) = G(g)$, then $T_f = T_g$ for all $x \in V$. This then says $T_f(v_i) = f(v_i) = g(v_i) = T_g(v_i)$ for all $v_i$, so that $f = g$ because they agree on all elements of $B$. It is surjective because $T \in \mathrm{Hom}_F(V,W)$ defines a map $f : B \to W$ by $f(v_i) = T(v_i)$ and by definition we have $G(f) = T$. It remains to show that $G$ is linear, however this is clear because $G(f+g) = T_{f+g} = T_f + T_g$ because $T_{f+g}(v_i) = (f+g)(v_i) = f(v_i) + g(v_i) = T_f(v_i) + T_g(v_i)$ so $T_{f+g}$ and $T_f + T_g$ agree on $B$ and therefore on all of $V$. Similarly we see $T_{cf} = cT_f$ for $c \in F$, so $G$ is linear and we are done. □

**Theorem 0.27.** *Two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.*

*Proof.* Suppose that $V, W$ are finite dimensional with $V \cong W$. Then by definition, there is a bijective linear transformation $T : V \to W$. By rank-nullity, $\mathrm{rank}(T) + \dim(\ker(T)) = \dim(V)$, and since $T$ is a bijection this says $\mathrm{Im}(T) = W$ so $\mathrm{rank}(T) = \dim(W)$ and $\dim(\ker(T)) = 0$, i.e. $\dim(V) = \dim(W)$. Now Suppose that $V$ and $W$ are vector spaces of the same dimension. Let $B = \{v_1,\ldots,v_n\}$ be a basis of $V$ and $B' = \{w_1,\ldots,w_n\}$ be a basis of $W$. Define $f : B \to W$ by $f(v_i) = w_i$. The previous theorem gives us a linear transformation $T_f : V \to W$. Write $x = c_1v_1+\ldots+c_nv_n$. Then if $T_f(x) = 0$, this says $c_1T(v_1)+\ldots+c_nT(v_n) = c_1w_1+\ldots+c_nw_n = 0$, so all $c_i = 0$ because $w_i$ are linearly independent. This says $x = 0$, so $\ker(T) = \{0\}$. Thus, $T$ is injective and therefore bijective, so $V \cong W$. □

## The matrix of a linear transformation

Throughout this section $V$ and $W$ are finite dimensional $F$-vector spaces of dimensions $n$ and $m$ respectively, with bases $\beta = \{v_1,\ldots,v_n\}$ and $\gamma = \{w_1,\ldots,w_m\}$.

For any vector $x \in V$, we may uniquely write $x = c_1v_1+\ldots+c_nv_n$, so the data of the vector $x$ is contained entirely in the list of coefficients of the basis vectors. This gives the following definition.

**Definition 0.28.** The **coordinate representation** of a vector $x = c_1v_1+\ldots+c_nv_n$ with respect to the basis $\beta = \{v_1,\ldots,v_n\}$ of $V$ is defined by $[x]_\beta = (c_1,\ldots,c_n) \in F^n$.

**Example 0.29.** Let $x = (1,3) \in \mathbb{R}^2$ and let $\beta = \{e_1,e_2\}$ be the standard basis of $\mathbb{R}^2$. Then $x = e_1 + 3e_2$ so $[x]_\beta = (1,3)$. Set $\gamma = \{(1,1),(1,-1)\}$. Then $x = 2(1,1) - (1,-1)$ so $[x]_\gamma = (2,-1)$. With $\alpha = \{(1,3),(1,0)\}$ we have $[x]_\alpha = (1,0)$.

**Example 0.30.** Let $\beta = \{1, x, x^2\}$ be the standard basis of $P_2(\mathbb{R})$. Then with $p(x) = 4 - 3x + 3x^2$, we have $[p(x)]_\beta = (4, -3, 3)$. If $\gamma = \{1, x, \frac{3}{2}x^2 - \frac{1}{2}\}$, we have $[p(x)]_\gamma = (5, -3, 2)$, as $5 - 3x + 2(\frac{3}{2}x^2 - \frac{1}{2}) = p(x)$.

**Example 0.31.** Let $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ be the standard basis of $M_2(\mathbb{R})$, and let $\gamma = \{E_{11}, E_{12} + E_{21}, E_{22}\}$ be a basis of $\mathrm{Sym}_2(\mathbb{R})$. Let $A = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$ Then viewed as an element of $M_2(F)$, we may write $[A]_\beta = (2, -1, -1, 5)$, but viewed as an element of $\mathrm{Sym}_2(\mathbb{R})$ we have $[A]_\gamma = (2, -1, 5)$.

**Example 0.32.** View $\mathbb{C}$ as an $\mathbb{R}$-vector space with basis $\beta = \{1, i\}$. Then $x = 3 + 5i$ has $[x]_\beta = (3, 5)$. As a $\mathbb{C}$-vector space, $\mathbb{C}$ has basis $\gamma = \{1\}$, so $[x]_\gamma = 3 + 5i$

**Proposition 6.** *The map $C_\beta : V \to F^n$ given by $C_\beta(x) = [x]_\beta$ gives an isomorphism $V \cong F^n$.*

*Proof.* Let $x, y \in V$ with $x = c_1v_1 + \ldots + c_nv_n$ and $y = d_1v_1 + \ldots + d_nv_n$. Then $x + y = (c_1 + d_1)v_1 + \ldots + (c_n + d_n)v_n$. We have $C_\beta(x + y) = (c_1 + d_1, \ldots, c_n + d_n) = (c_1, \ldots, c_n) + (d_1, \ldots, d_n) = C_\beta(x) + C_\beta(y)$. For any $k \in F$, we have $kx = kc_1v_1 + \ldots + kc_nv_n$, so $C_\beta(kx) = (kc_1, \ldots, kc_n) = k(c_1, \ldots, c_n) = kC_\beta(x)$. This proves that $C_\beta$ is linear. If $C_\beta(x) = 0$, this says that $x = 0v_1 + \ldots + 0v_n = 0$. This says $\ker(C_\beta) = \{0\}$, so $C_\beta$ is injective and therefore bijective giving $V \cong F^n$. □

Coordinates are one of the best ideas in mathematics, and in linear algebra this is no different. Coordinates give us a way of viewing a vector in an abstract vector space as a more concrete $n$-tuple of elements of $F$. In fact, we can do more: using coordinates, we can associate to every linear transformation $T : V \to W$ a matrix $[T]_\beta^\gamma \in M_{m\times n}(F)$. This reduces the study of linear maps from $V$ to $W$, and therefore linear algebra as a whole, to studying $M_{m\times n}(F)$.

For $x \in V$, write $x = c_1v_1 + \ldots + c_nv_n$, so $T(x) = c_1T(v_1) + \ldots + c_nT(v_n)$. Since the coordinate map $C_\gamma : W \to F^m$ is a linear transformation, this says $[T(x)]_\gamma = c_1[T(v_1)]_\gamma + \ldots + c_n[T(v_n)]_\gamma$. Set $[T(v_i)]_\gamma = (a_{1i}, \ldots, a_{mi})$. Written as a matrix equation, $[T(x)]_\gamma = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$.

**Definition 0.33.** Let $T : V \to W$ be a linear transformation. The **matrix of T** with respect to $\beta$ and $\gamma$ is denoted by $[T]_\beta^\gamma$ and is defined by $[T]_\beta^\gamma = \begin{pmatrix} | & & | \\ [T(v_1)]_\gamma & \ldots & [T(v_n)]_\gamma \\ | & & | \end{pmatrix}$. That is, $[T]_\beta^\gamma$ is the matrix whose columns are given by $[T(v_i)]_\gamma$. If $T : V \to V$ and $\beta = \gamma$, then we usually just write $[T]_\beta$.

The definition of the matrix of $T$ says that $[T(x)]_\gamma = [T]_\beta^\gamma[x]_\beta$, and so one can then recover the actual vector $T(x)$ by setting up the corresponding linear combination of basis vectors in $\gamma$.

**Example 0.34.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ with $T(x, y, z) = (x + 3z, -x + 2y + z, x + y + z)$. With $\beta = \{e_1, e_2, e_3\}$, and $\gamma = \{(1, -1, 1), (0, 2, 1), (3, 1, 1)\}$, we see $[T]_\beta = \begin{pmatrix} 1 & 0 & 3 \\ -1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, and $[T]_\beta^\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

**Example 0.35.** Let $T : M_2(F) \to M_2(F)$ be given by $T(A) = A - A^t$. With $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ the standard basis of $M_2(F)$, we have $[T]_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

**Example 0.36.** Let $\alpha = a + bi \in \mathbb{C}$. View $\mathbb{C}$ as an $\mathbb{R}$-vector space with the standard basis $\beta = \{1, i\}$ , and consider the linear transformation $T : \mathbb{C} \to \mathbb{C}$ defined by $T(x) = \alpha x$, the multiplication by $\alpha$ map. Then $[T]_\beta = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. This says any complex number $a + bi$ can be thought of as the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

**Example 0.37.** Let $D : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be the derivative map, and $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{1, x, x^2\}$ be the standard bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$. Then $[D]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$.

**Example 0.38.** Set $V = \text{Span}(\{\sin(x), \cos(x)\}) \subset C^\infty(\mathbb{R})$ and define $T : V \to V$ by $T(f) = 3f + 2f' - f''$. With $\beta = \{\sin(x), \cos(x)\}$, we see that $[T]_\beta = \begin{pmatrix} 4 & 4 \\ 2 & -2 \end{pmatrix}$ because $T(\sin(x)) = 4\sin(x) + 2\cos(x)$ and $T(\cos(x)) = 4\sin(x) - 2\cos(x)$. Using row reduction, one can check the only solution to $[T]_\beta x = 0$ is $x = 0$. This says no non-trivial linear combination of $\sin(x)$ and $\cos(x)$ are solutions the the differential equation $3f + 2f' - f'' = 0$.

**Example 0.39.** Let $T : V \to V$ be linear and suppose $W$ is a $T$-invariant subspace. Let $U$ be the complement of $W$ in $V$, so $V = W \oplus U$. If $\{w_1, \ldots, w_k\}$ and $\{u_1, \ldots, u_\ell\}$ are bases of $W$ and $U$, then $\{w_1, \ldots, w_k, u_1, \ldots, u_\ell\}$ is a basis of $V$. Since $T(W) \subset W$, we may write $T(w_i) = c_{i1}w_1 + \ldots + c_{ik}w_k$, so $[T(w_i)]_\beta = (c_{1i}, \ldots, c_{ki}, 0, \ldots, 0)$. This gives that $[T]_\beta$ is a block matrix of the form $\begin{pmatrix} A & B \\ O & C \end{pmatrix}$, where $A$ is the $k \times k$ matrix $[c_{ij}]$, $O$ is the $(n-k) \times (n-k)$ zero matrix, and $B$ and $C$ are some matrices of size $(n-k) \times (n-k)$ and $k \times k$ respectively. Therefore having a $T$-invariant subspace allows one to decompose the matrix of $T$ into an easier to work with block form.

The results of this section that there is a correspondence between matrices and linear transformations can be summed up in the below theorem.

**Lemma 0.40.** *Let $T, U : V \to W$ be linear transformations, and $c \in F$. Then $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$ and $[cT]_\beta^\gamma = c[T]_\beta^\gamma$.*

*Proof.* By definition, the $i$-th column of $[T+U]_\beta^\gamma$ is equal to $[(T+U)(v_i)]_\gamma = [T(v_i)+U(v_i)]_\gamma = [T(v_i)]_\gamma + [U(v_i)]_\gamma$ by linearity of the map $C_\gamma$. However, this is clearly also the $i$-th column of the matrix $[T]_\beta^\gamma + [U]_\beta^\gamma$ so $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$. Similarly, the $i$-th column of the matrix $[cT]_\beta^\gamma$ is given by $[(cT)(v_i)]_\gamma = [cT(v_i)]_\gamma = c[T(v_i)]_\gamma$, which is again the $i$-th column of the matrix $c[T]_\beta^\gamma$. □

**Theorem 0.41.** *$Hom_F(V,W) \cong M_{m\times n}(F)$. So in particular,* $\dim(Hom_F(V,W)) = mn$.

*Proof.* Define $F : \mathrm{Hom}_F(V,W) \to M_{m\times n}(F)$ by $F(T) = [T]_\beta^\gamma$. Suppose that $F(T) = F(U)$. Then $[T]_\beta^\gamma = [U]_\beta^\gamma$, so in particular the columns of these matrices are the same so $[T(v_i)]_\gamma = [U(v_i)]_\gamma$ for all $i$. Translating back to the actual vectors says $T(v_i) = U(v_i)$, i.e. $T = U$ so $F$ is injective. For a matrix $A = \begin{pmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{pmatrix} \in M_{m\times n}(F)$, write $x_i = (a_{1i}, \dots, a_{mi})$. Then define $f : B \to W$ by $f(v_i) = a_{1i}w_1 + \ldots + a_{mi}w_m$. This defines a linear transformation $T : V \to W$ with $T(v_i) = f(v_i)$ so in coordinates, $[T(v_i)]_\gamma = [f(v_i)]_\gamma = x_i$. This then says $[T]_\beta^\gamma = A$, so that $F$ is surjective, so $F$ is a bijection. By the above lemma, $F$ is linear, so $F$ is an isomorphism as desired. The dimension result then follows immediately. □

## Invertibility

Recall that a function $f : X \to Y$ is said to be invertible if there is $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$, and we denote $g = f^{-1}$. From set theory, $f$ is invertible if and only if $f$ is a bijection. For linear transformations $T : V \to W$ and $S : W \to Z$, we denote the composition $S \circ T$ by $ST$, and clearly then $T$ is an isomorphism if and only if $T$ is invertible. Since linear transformations correspond to matrices, we make a similar definition.

**Definition 0.42.** Let $A \in M_n(F)$. Then $A$ is **invertible** if there is $B \in M_n(F)$ such that $AB = BA = I_n$. If such a matrix exists it's easy to see that it must be unique, and we then write $B = A^{-1}$.

**Definition 0.43.** Let $V$ be a vector space. We define $\mathrm{GL}(V) = \{T \in \mathrm{End}_F(V) : T \text{ is invertible}\}$. Similarly, we set $\mathrm{GL}_n(F) = \{A \in M_n(F) : A \text{ is invertible}\}$.

**Proposition 7.** *Let $T : V \to W$ and $S : W \to Z$ be linear transformations. Then $ST : V \to Z$ is a linear transformation. If $T$ is invertible, then $T^{-1}$ is a linear transformation.*

*Proof.* For $x, y \in V$ we have $ST(x+y) = S(T(x+y)) = S(T(x)+T(y)) = ST(x)+ST(y)$, and for $c \in F$, we also have $ST(cx) = S(T(cx)) = S(cT(x)) = cST(x)$ since $S, T$ are linear. Suppose that $T$ is invertible with inverse $T^{-1}$. For $w, w' \in W$, $T^{-1}(w+w')$ is the vector that maps to $w+w'$ under $T$. Since $T$ is linear, $T(T^{-1}(w)+T^{-1}(w')) = w+w'$, so $T^{-1}(w+w') = T^{-1}(w)+T^{-1}(w')$. Similarly we see for $c \in F$ that $T^{-1}(cw) = cT^{-1}(w)$ so $T^{-1}$ is linear.

□

**Proposition 8.** *Let $S : W \to Z$ and $T : V \to W$ be linear transformations, and let $\alpha, \beta, \gamma$ be bases of the finite dimensional vector spaces $V, W, Z$ respectively. Then $[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta$*

*Proof.* Let $\alpha = \{v_1, \dots, v_n\}$ and $\beta = \{w_1, \dots, w_k\}$. By definition, the $i$-th column of $[ST]_\alpha^\gamma$ is given by $[ST(v_i)]_\gamma$. The $i$-th column of $[S]_\beta^\gamma [T]_\alpha^\beta$ is $S_\beta^\gamma [T(v_i)]_\beta$, so it's sufficient to check these expressions are equal. Write $T(v_i) = c_{1i}w_1 + \ldots + c_{ki}w_k$. Then $ST(v_i) = S(c_{1i}w_1 +$

$\ldots + c_{ki}w_k) = c_{1i}S(w_1) + \ldots + c_{1k}S(w_k)$. Applying $C_\gamma$ then gives $[ST(v_i)]_\gamma = c_{1i}[S(w_1)]_\gamma + \ldots + c_{1k}[S(w_k)]_\gamma$, which we then recognize as saying $[ST(v_i)]_\gamma = [S]_\beta^\gamma[T(v_i)]_\beta$ as desired. □

**Proposition 9.** *Let $T : V \to W$ be a linear transformation, and let $\beta, \gamma$ be bases of the finite dimensional vector spaces $V$ and $W$. Then $T$ is invertible if and only if $[T]_\beta^\gamma$ is invertible, and further $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$.*

*Proof.* If $T$ is invertible, then $T^{-1} : W \to V$ satisfies $TT^{-1} = \mathrm{id}_W$ and $TT^{-1} = \mathrm{id}_V$, so the above says $[T]_\beta^\gamma[T^{-1}]_\gamma^\beta = [T^{-1}]_\gamma^\beta[T]_\beta^\gamma = I_n$, so that $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$. Conversely, suppose that $[T]_\beta^\gamma$ is invertible. Then the columns of $[T]_\beta^\gamma$ are linearly independent: if not, there are $c_1, \ldots, c_n \in F$ not all 0 such that $c_1[T(v_1)]_\gamma + \ldots + c_n[T(v_n)]_\gamma = 0$, i.e. there is a non-trivial solution to $[T]_\beta^\gamma x = 0$. However, this is impossible because multiplying by the inverse of $[T]_\beta^\gamma$ on the left shows that if the above holds then necessarily $x = 0$. This says that the vectors $[T(v_i)]_\gamma$ in $F^n$ are linearly independent, and as the coordinate mapping is an isomorphism this then implies that the vectors $w_i = T(v_i)$ are linearly independent vectors in $W$, and therefore are a basis of $W$. Define a linear transformation $S : W \to V$ by $S(w_i) = v_i$ and extend linearly. By definition, $ST(v_i) = S(w_i) = v_i$, so $ST = \mathrm{id}_V$, and similarly $TS(w_i) = T(v_i) = w_i$, so $TS = \mathrm{id}_W$ so that $T$ is invertible as desired. □

Knowing if a linear transformation is an isomorphism or not is extremely important – if two vector spaces $V$ and $W$ are isomorphic, this essentially says that $W$ and $V$ are the "same" vector space up to relabeling of the elements, because addition of vectors in one space corresponds uniquely to addition of vectors in the other. This then reduces the study of vector spaces to studying vector spaces up to isomorphism. If $T$ is a linear operator on some vector space $V$, knowing that $T$ is invertible is very powerful, as will hopefully be demonstrated in the following examples.

**Example 0.44.** Let $V = \mathrm{Span}(\{e^{ax}\sin(bx), e^{ax}\cos(bx)\}) \subset C^\infty(\mathbb{R})$ for $a, b \neq 0$, and let $D$ be the differential operator. Then $V$ is a 2-dimensional $D$-invariant subspace. With $\beta = \{e^{ax}\sin(bx), e^{ax}\cos(bx)\}$, the matrix $[D|_V]_\beta$ is given by $[D|_V]_\beta = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Since $\det([D|_V]_\beta) = a^2 + b^2 \neq 0$, $D|_V$ is invertible with inverse $A = \begin{pmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ -\frac{b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{pmatrix}$. As $\int f'(x)\,dx = f(x) + C$ and $\frac{d}{dx}\int f(x)\,dx = f(x)$, choosing the choice of constant to be $C = 0$ says the inverse of $D$ is the indefinite integral operator. To integrate say, $\int e^{ax}\sin(bx)\,dx$, we see $A[e^{ax}\sin(bx)]_\beta = (\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2})$, so converting back into vectors of $V$ says $\int e^{ax}\sin(bx)\,dx = \frac{a}{a^2+b^2}e^{ax}\sin(bx) - \frac{b}{a^2+b^2}e^{ax}\cos(bx) + C$ after appending back the arbitrary constant of integration.

**Example 0.45.** ASCII is an encoding standard that associates characters to 7-digit binary strings, which we may think of as elements of $(\mathbb{Z}/2\mathbb{Z})^7$. Fix a matrix $A \in \mathrm{GL}_7(\mathbb{Z}/2\mathbb{Z})$. A simple encryption method is as follows: given a message $M$, convert each character $c$ of $M$ to ASCII and then convert it into a vector $x_c \in (\mathbb{Z}/2\mathbb{Z})^7$. Encrypt $M$ character-wise by computing $Ax_c$ for all characters, and convert back to text. Since $M$ is invertible, the message can be decrpyted by again converting text to ASCII and multiplying characters by $A^{-1}$. As an example, the message "TEST" corresponds to the block of binary strings

"1010100 1000101 1010011 1010100". With $A = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$, this gets encrypted to the message "gS∼g", which can be decrypted if $A$ or $A^{-1}$ is known.

We showed above that checking the invertibility of a linear operator $T$ is the same as checking the invertibility of its corresponding matrix. We remind the reader of some of many equivalent conditions for checking the latter:

**Theorem 0.46.** *Let $A \in M_n(F)$. Then the following are equivalent:*

(a) *$A$ is invertible.*
(b) *The only solution in $F^n$ to $Ax = 0$ is $x = 0$.*
(c) *The columns of $A$ are linearly independent.*
(d) *$A$ is row-equivalent to $I_n$.*
(e) $\det(A) \neq 0$.
(f) *The augmented matrix $[A|I]$ is row equivalent to $[I|B]$ for some non-zero matrix $B$.*

## Change of basis and similarity

Given a linear operator $T$ on $V$, the matrix $[T]_\beta$ depends on a choice of basis $\beta$ of $V$. Picking a different basis $\beta'$ will produce a different looking matrix $[T]_{\beta'}$, but it still represents the same operator $T$. A natural question is given two matrices $A, B \in M_n(F)$, how can one check if they come from the same linear operator in $\mathrm{GL}(V)$?

Pick bases $\beta, \gamma$ of $V$, and consider the identity operator $\mathrm{id}_\mathrm{V}$, along with the corresponding matrix $[\mathrm{id}_V]_\beta^\gamma$. This matrix satisfies $[x]_\gamma = [\mathrm{id}_V(x)]_\gamma = [\mathrm{id}_V]_\beta^\gamma [x]_\beta$ for all $x \in V$, or in other words, multiplication by $[\mathrm{id}_V]_\beta^\gamma$ converts the coordinates of the vector $x$ from the basis $\beta$ to the basis $\gamma$.

**Definition 0.47.** Let $V$ be a vector spaces with basis $\beta = \{v_1, \dots, v_n\}$, and let $\gamma = \{v_1', \dots, v_n'\}$ be another basis. The **change of basis matrix** from $\beta$ to $\gamma$, denoted $S_\beta^\gamma$, is the matrix $[\mathrm{id}_V]_\beta^\gamma$. Explicitly, $S_\beta^\gamma = \begin{pmatrix} | & & | \\ [v_1]_\gamma & \dots & [v_n]_\gamma \\ | & & | \end{pmatrix}$.

Since $\mathrm{id}_V$ is invertible, this says $S_\beta^\gamma$ is invertible, and has inverse matrix $[\mathrm{id}_V]_\gamma^\beta = S_\gamma^\beta$.

**Theorem 0.48.** *Let $\beta, \gamma$ be two bases of a finite dimensional vector space $V$, and let $T \in GL(V)$. Then $[T]_\gamma = S_\beta^\gamma [T]_\beta S_\gamma^\beta$, and $[T]_\gamma S_\beta^\gamma = S_\beta^\gamma [T]_\beta$. In otherwords, the following diagram*

$$\begin{tikzcd} {[x]_\beta} \arrow[r, "{[T]_\beta}"] \arrow[d, "S_\beta^\gamma"] & {[T(x)]_\beta} \arrow[d, "S_\beta^\gamma"] \\ {[x]_\gamma} \arrow[r, "{[T]_\gamma}"] & {[T(x)]_\gamma} \end{tikzcd}$$

*commutes.*

*Proof.* Since composition of linear transformations corresponds to multiplication by their corresponding matrices, we see $[T]_\gamma = [\mathrm{id}_V \circ T \circ \mathrm{id}_V]_\gamma^\gamma = [\mathrm{id}_V]_\beta^\gamma [T]_\beta [\mathrm{id}_V]_\gamma^\beta = S_\beta^\gamma [T]_\beta S_\gamma^\beta$. Since $S_\beta^\gamma$ is invertible with inverse $S_\gamma^\beta$, multiplication on the left gives $S_\gamma^\beta [T]_\gamma = [T]_\beta S_\gamma^\beta$ as desired. ☐

**Example 0.49.** Let $\beta = \{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$ and $\gamma = \{(1,1),(1,2)\}$ be another basis. The change of basis matrix $S_\gamma^\beta$ is $S_\gamma^\beta = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. To compute $S_\beta^\gamma$, we take the inverse to find $S_\beta^\gamma = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. To compute $[e_1]_\gamma$, we see $[e_1]_\gamma = S_\beta^\gamma [e_1]_\beta = S_\beta^\gamma e_1 = (2,-1)$, so that $(1,0) = 2(1,1) - (1,2)$.

**Example 0.50.** Let $\beta = \{1, x, x^2\}$ and $\gamma = \{1, x, \frac{3}{2}x^2 - \frac{1}{2}\}$ be bases of $P_2(\mathbb{R})$. Then $S_\gamma^\beta = \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{pmatrix}$, and one can compute $S_\beta^\gamma = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}$. Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $T(f)(x) = xf'(x)$. Then $[T]_\beta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and the change of basis formula says $[T]_\gamma = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

**Example 0.51.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x,y,z) = (2z, -2x+3y+2z, -x+3z)$. Let $\beta = \{e_1, e_2, e_3\}$ be the standard basis of $\mathbb{R}^3$, and let $\gamma = \{(2,1,1),(1,0,1),(0,1,0)\}$ be another basis. Then $[T]_\beta = \begin{pmatrix} 0 & 0 & 2 \\ -2 & 3 & 2 \\ -1 & 0 & 3 \end{pmatrix}$. We see $S_\gamma^\beta = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, and $S_\beta^\gamma = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix}$, so the change of basis formula says $[T]_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. With respect to the new basis $\gamma$, this says that $T$ acts along each $\gamma$-direction by scaling. Having a basis where an operator is diagonal is extremely useful, as it allows one to easily compute values of compositions. For example, to compute $T^n(1,2,3)$, we compute $[T^n(1,2,3)]_\gamma = [T^n]_\gamma [(1,2,3)]_\gamma = [T]_\gamma^n [(1,2,3)]_\gamma$. We have $[(1,2,3)]_\gamma = S_\beta^\gamma [(1,2,3)]_\beta = (-2,5,4)$, so $[T^n(1,2,3)]_\gamma = [T]_\gamma^n (-2,5,4)^T = (-2, 5\cdot 2^n, 4\cdot 3^n)$. This says $T^n(1,2,3) = -2(2,1,1) + 5\cdot 2^n(1,0,1) + 4\cdot 3^n(0,1,0) = (-4+5\cdot 2^n, -2+4\cdot 3^n, -2+5\cdot 2^n)$.

**Example 0.52.** Consider $P_3(\mathbb{R})$, and set $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{\binom{x}{0}, \ldots, \binom{x}{3}\}$, where $\binom{x}{0} = 1$ and for $k \geq 1$ we have $\binom{x}{k} = \frac{x(x-1)\ldots(x-k+1)}{k!}$. Then $\gamma$ is a basis, because each polynomial in $\gamma$ has a different degree. The change of basis matrix $S_\gamma^\beta$ is given by $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{6} \end{pmatrix}$, and one can check that $S_\beta^\gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 6 \end{pmatrix}$. Then $[x^3]_\gamma = S_\beta^\gamma [x^3]_\beta = (0,1,6,6)$, so $x^3 = \binom{x}{1} + 6\binom{x}{2} + 6\binom{x}{3}$. As an application, $\sum_{k=1}^{n-1} k^3 = \sum_{k=1}^{n-1} \binom{k}{1} + 6\binom{k}{2} + 6\binom{k}{3} = \binom{n}{2} + 6\binom{n}{3} + 6\binom{n}{4} = (\frac{n(n-1)}{2})^2$, which follows from the identity $\sum_{k=1}^{n-1} \binom{k}{r} = \binom{n}{r+1}$, easily proven by induction.

The theorem from above leads us to the following definition:

**Definition 0.53.** For $A, B \in M_n(F)$, we say that $A$ and $B$ are **similar** and write $A \sim B$ if there exists $P \in \mathrm{GL}_n(F)$ such that $A = PBP^{-1}$.

Since $S_\gamma^\beta = (S_\beta^\gamma)^{-1}$, this says that for any choice of bases $\beta, \gamma$ of $V$ the matrices $[T]_\beta$ and $[T]_\gamma$ are similar. The following observation is easy to verify:

**Proposition 10.** *Similarly is an equivalence relation on $M_n(F)$.*

We showed that changing the basis of $V$ from $\beta$ to $\gamma$ yields similar matrices $[T]_\beta$ and $[T]_\gamma$. The converse is true as well:

**Theorem 0.54.** *$A \sim B$ in $M_n(F)$ if and only if there are bases $\beta, \gamma$ of $V$ and a linear transformation $T$ such that $A = [T]_\beta$ and $B = [T]_\gamma$. That is, similar matrices correspond to the same linear transformation under potentially different bases.*

*Proof.* Suppose that $A \sim B$ in $M_n(F)$, so there is $P \in \mathrm{GL}_n(F)$ such that $A = PBP^{-1}$, and let $\beta = \{v_1, \ldots, v_n\}$ be any basis of $V$. We have seen that we may choose $T$ such that $[T]_\beta = A$, so it remains to find $\gamma$ such that $[T]_\gamma = B$. To do this, we would like to think of $P$ as some change of basis matrix. Define $w_i$ such that $[w_i]_\beta = Pe_i$, where $e_i$ are the standard basis vectors of $F^n$. As $P$ is invertible, its columns are linearly independent, and because the coordinate map $C_\beta$ is an isomorphism, $w_i$ are also linearly independent so that $\gamma = \{w_1, \ldots, w_n\}$ is a basis of $V$. Then $S_\gamma^\beta$ is the matrix with columns $[w_i]_\beta = Pe_i$, so $S_\gamma^\beta = P$, and $S_\beta^\gamma = P^{-1}$. Since $A = PBP^{-1}$, this says $B = P^{-1}AP = S_\beta^\gamma [T]_\beta S_\gamma^\beta = [T]_\gamma$ as desired. The backwards direction was proven above.

$\square$

The conjugacy classes of matrices in $M_n(F)$ under similarly correspond to the *distinct* linear operators $T \in \mathrm{GL}(V)$, regardless of choice of basis. Therefore if one cares only about the different types of operators that arise on $V$, the importance of studying matrices up to similarity is self evident. To be able to distinguish between conjugacy classes, it's helpful to known some quantities that are invariant under similarity.

**Lemma 0.55.** *Let $A, B \in M_n(F)$. Then $tr(AB) = tr(BA)$.*

*Proof.* Write $A = (a_{ij})$ and $B = (b_{ij})$. Then $\mathrm{tr}(AB) = \sum_{i=1}^n (AB)_{ii} = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{ki}$. On the other hand, $\mathrm{tr}(BA) = \sum_{i=1}^n (BA)_{ii} = \sum_{i=1}^n \sum_{k=1}^n b_{ik} a_{ki}$. By renaming variables $i$ and $k$, we have $\sum_{i=1}^n \sum_{k=1}^n b_{ik} a_{ki} = \sum_{k=1}^n \sum_{i=1}^n b_{ki} a_{ik}$, and by swapping the order of summation this equals $\sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{ki}$ as desired. $\square$

**Proposition 11.** *Let $A, B \in M_n(F)$, with $A \sim B$. Then $tr(A) = tr(B)$ and $det(A) = det(B)$.*

*Proof.* As $A \sim B$, there is $P \in M_n(F)$ such that $A = PBP^{-1}$. By the lemma, $\operatorname{tr}(A) = \operatorname{tr}(P(BP^{-1})) = \operatorname{tr}((BP^{-1})P) = \operatorname{tr}(B)$. Similarly, we find that $\det(A) = \det(PBP^{-1}) = \det(P)\det(B)\det(P^{-1}) = \det(B)\det(P)\det(P^{-1}) = \det(B)\det(I_n) = \det(B)$ by properties of the determinant. □

Since similarity corresponds to change of basis, this allows us to define these quantities for a linear operator.

**Definition 0.56.** Let $V$ be finite dimensional. For $T \in \mathrm{GL}(V)$ we define the **trace** of $T$ and the **determinant** of $T$ to be the quantities $\operatorname{tr}([T]_\beta)$ and $\det([T]_\beta)$ for any choice of basis $\beta$ of $V$.

**Example 0.57.** The matrices $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 2 \\ 1 & -3 \end{pmatrix}$ are not similar, because $\operatorname{tr}(A) = 5$ while $\operatorname{tr}(B) = -3$. However, both $\det(A) = \det(B) = -2$.

Recall that the rank of a linear transformation $T$ was defined as the dimension of it's image. We can also similarly define rank in terms of a matrix representation of $T$:

**Proposition 12.** *Let $A, B \in M_n(F)$ and suppose that $A \sim B$. Then $\operatorname{rank}(A) = \operatorname{rank}(B)$.*

*Proof.* Write $A = PBP^{-1}$ for some $P$. Define $T : \operatorname{Im}(A) \to F^n$ by $T(x) = P^{-1}x$. Then $T$ is injective because $P$ is invertible, so $\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(T))$. We also see that if $y \in \operatorname{Im}(T)$, then $y = P^{-1}x$ for some $x \in \operatorname{Im}(A)$. Write $x = Az$, so $y = P^{-1}(Az) = B(P^{-1}z) \in \operatorname{Im}(B)$. This gives $\operatorname{rank}(A) \leq \operatorname{rank}(B)$. Similarly with $S : \operatorname{Im}(B) \to F^n$ defined by $S(x) = Px$, we find $\operatorname{rank}(B) \leq \operatorname{rank}(A)$, so that $\operatorname{rank}(A) = \operatorname{rank}(B)$. □