NOTES ON LINEAR TRANSFORMATIONS

Basic properties

Definition 1. Let V and W be vector spaces. A function $T\colon V \to W$ is a *linear transformation from V to W* if the following two properties hold.

(i) $T(v + v') = T(v) + T(v')$ for all $v, v' \in V$.
(ii) $T(av) = aT(v)$ for all $v \in V$ and all scalars $a \in F$.

It is important to note that the addition of vectors $v + v'$ in (i) takes place in V, whereas the addition $T(v) + T(v')$ takes place in W. Similarly, av in (ii) is scalar multiplication in V, whereas $aT(v)$ is scalar multiplication in W. It is common to refer to T as a *structure preserving* map, because it respects the two basic operations that give V and W their vector space structures, namely addition and scalar multiplication.

Example 2.

(a) Let V be any vector space and consider the *identity function* $\mathrm{id}_V\colon V \to V$ given by $\mathrm{id}_V(v) = v$ for all $v \in V$. To see that id_V is a linear transformation, observe for all $v, v' \in V$ that we have

$$\mathrm{id}_V(v + v') = v + v' = \mathrm{id}_V(v) + \mathrm{id}_V(v'),$$

hence the first condition in Definition 1 holds for id_V. Moreover if $v \in V$, and $a \in F$ is any scalar, then we have

$$\mathrm{id}_V(av) = av = a\,\mathrm{id}_V(v).$$

This shows that the second condition in Definition 1 holds for the function id_V, thus id_V is a linear transformation.

(b) Let V and W be arbitrary vector spaces and consider the *zero function* $Z\colon V \to W$ given by $Z(v) = 0_W$ for all $v \in V$. We have

$$Z(v + v') = 0_W = 0_W + 0_W = Z(v) + Z(v')$$

and

$$Z(av) = 0_W = a0_W = aZ(v)$$

for all $v, v' \in V$ and $a \in F$. This shows that Z is a linear transformation.

(c) Consider the function $\pi\colon \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\pi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

This is called the *projection onto the x-y-plane*. To show that π is a linear transformation, observe for any $x, y, z, x', y', z' \in \mathbb{R}$ that we have

$$\pi\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}\right) = \pi\begin{pmatrix} x+x' \\ y+y' \\ z+z' \end{pmatrix} = \begin{pmatrix} x+x' \\ y+y' \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} x' \\ y' \\ 0 \end{pmatrix}$$
$$= \pi\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \pi\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}.$$

For any $x, y, z, a \in \mathbb{R}$ we also have

$$\pi\left(a\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \pi\begin{pmatrix} ax \\ ay \\ az \end{pmatrix} = \begin{pmatrix} ax \\ ay \\ 0 \end{pmatrix} = a\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = a\,\pi\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

These relations show that π is a linear transformation.

(d) Given a fixed real number θ, let $\rho_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote the function that takes a point in the x-y-plane and rotates it by θ radians counter clockwise about the origin. It is a standard exercise to show that ρ_θ is a linear transformation.

(e) Let $\sigma\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the function that reflects a point in the x-y-plane about the line $y = x$. Specifically,

$$\sigma\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

It is straightforward to verify that σ is a linear transformation. More generally, any reflection of the x-y-plane about a line that passes through the origin will be a linear transformation.

(f) Let $\mathrm{Fun}(\mathbb{R}, \mathbb{R})$ be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$ and let $\mathrm{Dif}(\mathbb{R}, \mathbb{R})$ be the set of all such functions that are differentiable. Recall that both of these sets are $\mathbb{R}$-vector spaces. Taking the derivative

$$\tfrac{d}{dx}\colon \mathrm{Dif}(\mathbb{R}, \mathbb{R}) \longrightarrow \mathrm{Fun}(\mathbb{R}, \mathbb{R})$$

is a linear transformation. Indeed, it is well known from calculus that if f and g are differentiable functions and $c \in \mathbb{R}$ is a constant, then

$$\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$$

and

$$\frac{d}{dx}(cf) = c\frac{df}{dx}.$$

Similarly, if $\mathrm{Con}(\mathbb{R},\mathbb{R})$ is the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$, then the function

$$\textstyle\int\colon \mathrm{Con}(\mathbb{R},\mathbb{R}) \longrightarrow \mathrm{Dif}(\mathbb{R},\mathbb{R})$$

given by

$$\textstyle\int(f) = \displaystyle\int_0^x f(t)\,dt$$

is a linear transformation. Indeed, it is well known from calculus that if f and g are continuous functions and $c \in \mathbb{R}$ is a constant, then

$$\int_0^x (f(t)+g(t))\,dt = \int_0^x f(t)\,dt \,+\, \int_0^x g(t)\,dt$$

and

$$\int_0^x cf(t)\,dt = c\int_0^x f(t)\,dt.$$

(g) If F is any field, let $\mathcal{P}_n(F)$ denote the set of all polynomials of degree at most n in the variable x. Note that the concept of 'limit' no longer makes sense in this context like it did in calculus, but we can still define something called the *formal derivative*, which is the function

$$\tfrac{d}{dx}\colon \mathcal{P}_n(F) \longrightarrow \mathcal{P}_{n-1}(F)$$

given by

$$\tfrac{d}{dx}\big(a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0\big) = na_nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2x + a_1.$$

It is worthwhile to confirm that the formal derivative is a linear transformation.

Now let F be the field $\mathbb{Q}$ of rational numbers. We define the *formal integral* to be the function

$$\textstyle\int\colon \mathcal{P}_{n-1}(\mathbb{Q}) \longrightarrow \mathcal{P}_n(\mathbb{Q})$$

given by

$$\textstyle\int\big(b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_1x + b_0\big) = \dfrac{b_{n-1}}{n}x^n + \dfrac{b_{n-2}}{n-1}x^{n-1} + \cdots + \dfrac{b_1}{2}x^2 + b_0x.$$

One may also confirm that the formal integral is a linear transformation. Note that the choice $F = \mathbb{Q}$ was necessary because dividing by certain integers might not make sense in an arbitrary field. For example, the prime number p is not invertible in the field $\mathbb{Z}/p$, so the formula for the formal integral would not be legal from $\mathcal{P}_{p-1}(\mathbb{Z}/p)$ to $\mathcal{P}_p(\mathbb{Z}/p)$.

The following proposition allows us take some of the above examples and form new linear transformations.

Lemma 3. *If $S\colon U \to V$ and $T\colon V \to W$ are linear transformations, then their composition $T \circ S\colon U \to W$ is also a linear transformation.*

Proof. This is left as a standard exercise. $\square$

Example 4. Let $\sigma_x\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $\sigma_y\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the reflections of the x-y-plane about the x-axis and y-axis, respectively. These are both linear transformations by Example 2(e). One may check that their composition (in either order!) is none other than the reflection of the x-y-plane about the origin, hence the latter is also a linear transformation.

Proposition 5. *Let V and W be vector spaces and let $T\colon V \to W$ be a linear transformation from V to W. Then $T(0_V) = 0_W$, where 0_V and 0_W are the zero vectors in V and W, respectively.*

Proof. Since T is a linear transformation, we have

$$T(0_V) = T(0_V + 0_V) = T(0_V) + T(0_V).$$

Adding the additive inverse of $T(0_V)$ in W to both sides yields $0_W = T(0_V)$. $\square$

Linear transformations give rise to two important examples of subspaces.

Definition 6. Let V and W be any vector spaces and let $T\colon V \to W$ be a linear transformation from V to W. The *kernel* of T is defined to be the subset

$$\mathrm{Ker}(T) = \{v \in V \mid T(v) = 0_W\}.$$

of V. The *image* of T is defined to be the subset

$$\mathrm{Im}(T) = \{w \in W \mid \text{there exists } v \in V \text{ such that } w = T(v)\}$$

of W. Note that the image of T is simply a fancy name for the range of the function T.

Proposition 7. *If V and W are vector spaces and $T\colon V \to W$ is a linear transformation, then $\mathrm{Ker}(T)$ and $\mathrm{Im}(T)$ are subspaces of V and W, respectively.*

Proof. Choose $v, v' \in \mathrm{Ker}(T)$. Then $T(v) = 0_W$ and $T(v') = 0_W$ so that

$$T(v + v') = T(v) + T(v') = 0_W + 0_W = 0_W.$$

This shows that $v + v' \in \mathrm{Ker}(T)$, so $\mathrm{Ker}(T)$ is closed under addition. Now let $v \in \mathrm{Ker}(T)$ and $a \in F$. Then

$$T(av) = aT(v) = a0_W = 0_W.$$

This shows that $av \in \mathrm{Ker}(T)$, so $\mathrm{Ker}(T)$ is also closed under scalar multiplication. Finally, by Proposition 5, we have $T(0_V) = 0_W$. This shows that $0_V \in \mathrm{Ker}(T)$. These results imply that $\mathrm{Ker}(T)$ is a subspace of V.

The proof that $\mathrm{Im}(T)$ is a subspace of W is left to the reader. $\square$

Example 8. One may check that the kernel of the linear transformation π in Example 2(c) is the set of all vectors having zero x- and y-coordinates, which is interpreted geometrically as the z-axis. The image of π is the x-y-plane.

Example 9. The kernel and image of the formal derivative $\frac{d}{dx}\colon \mathcal{P}_n(F) \to \mathcal{P}_{n-1}(F)$ depend on the field F. For example, if $F = \mathbb{Q}$, then for any $n \geq 1$ the only polynomials annihilated by differentiation are the constants, so in this case we have $\operatorname{Ker}\left(\frac{d}{dx}\right) = \operatorname{Span}\{1\}$. Also, every polynomial of degree at most $n-1$ is the derivative of some polynomial in $\mathcal{P}_n(\mathbb{Q})$ (take the integral), so $\operatorname{Im}\left(\frac{d}{dx}\right) = \mathcal{P}_{n-1}(\mathbb{Q})$.

Now let $n = p$ be a prime number and $F = \mathbb{Z}/p$. We then have

$$\begin{aligned}\tfrac{d}{dx}\left(a_p x^p + a_{p-1}x^{p-1} + \cdots + a_1 x + a_0\right) &= p a_p x^{p-1} + (p-1)a_{p-1}x^{p-2} + \cdots + 2a_2 x + a_1 \\ &= (p-1)a_{p-1}x^{p-2} + \cdots + 2a_2 x + a_1\end{aligned}$$

since $p = 0$ in $\mathbb{Z}/p$. It follows that $\operatorname{Ker}\left(\frac{d}{dx}\right) = \operatorname{Span}\{x^p, 1\}$. We can also see from this that any polynomial in the image of differentiation will not involve the monomial x^{p-1}, hence

$$\operatorname{Im}\left(\tfrac{d}{dx}\right) = \operatorname{Span}\left\{x^{p-2}, \ldots, x, 1\right\}.$$

Definition 10. If V and W are finite dimensional vector spaces and $T\colon V \to W$ is a linear transformation, then we know that $\operatorname{Ker}(T)$ and $\operatorname{Im}(T)$ are also finite dimensional. We define the *nullity* of T to be the dimension of $\operatorname{Ker}(T)$. Similarly, we define the *rank* of T to be the dimension of $\operatorname{Im}(T)$. For convenience, we denote the nullity of T by $\operatorname{nullity}(T)$ and the rank of T by $\operatorname{rank}(T)$.

The rank and nullity of a linear transformation satisfy a very interesting relationship.

Theorem 11 (Rank-nullity theorem)**.** *Let V and W be finite dimensional vector spaces and let $T\colon V \to W$ be a linear transformation. Then*

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim V.$$

Proof. We choose a basis $u_1, \ldots, u_m$ of $\operatorname{Ker}(T)$ and use the replacement theorem to extend this to a basis $u_1, \ldots, u_m, v_1, \ldots, v_r$ of V. We have $\operatorname{nullity}(T) = m$ and $r + m = \dim V$. To prove the theorem, it therefore suffices to show that $\operatorname{rank}(T) = r$. To this end, we wish to prove that the vectors $T(v_1), \ldots, T(v_r)$ are a basis of $\operatorname{Im}(T)$.

To show that these vectors span $\operatorname{Im}(T)$, note that any vector in V may be written as a linear combination

$$a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_r v_r.$$

It follows that any element w of $\operatorname{Im}(T)$ is of the form

$$w = T(a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_r v_r).$$

But the right hand side is equal to

$$T(a_1u_1 + \cdots + a_mu_m) + b_1T(v_1) + \cdots + b_rT(v_r)$$

since T is a linear transformation. Observe that $a_1u_1 + \cdots + a_mu_m$ lies in $\mathrm{Ker}(T)$, because $u_1, \ldots, u_m$ is a basis of $\mathrm{Ker}(T)$. It follows that $T(a_1u_1 + \cdots + a_mu_m) = 0_W$, giving us

$$w = b_1T(v_1) + \cdots + b_rT(v_r).$$

This shows that every vector in $\mathrm{Im}(T)$ is a linear combination of $T(v_1), \ldots, T(v_r)$, so these vectors span $\mathrm{Im}(T)$.

To show that $T(v_1), \ldots, T(v_r)$ are linearly independent, suppose that

$$b_1T(v_1) + \cdots + b_rT(v_r) = 0_W.$$

We need to show that $b_1 = \cdots = b_r = 0$. Using the fact that T is a linear transformation, the above equality becomes

$$T(b_1v_1 + \cdots + b_rv_r) = 0_W.$$

This implies that $b_1v_1 + \cdots + b_rv_r$ lies in $\mathrm{Ker}(T)$. Because $u_1, \ldots, u_m$ is a basis of $\mathrm{Ker}(T)$, we may write $b_1v_1 + \cdots + b_rv_r$ as a linear combination

$$b_1v_1 + \cdots + b_rv_r = a_1u_1 + \cdots + a_mu_m$$

for some coefficients $a_1, \ldots, a_m$. But then

$$-a_1u_1 - \cdots - a_mu_m + b_1v_1 + \cdots + b_rv_r = 0_V.$$

Notice that the collection $u_1, \ldots, u_m, v_1, \ldots, v_r$, being a basis of V, is linearly independent. This means that all of the coefficients in the above expression must be zero. In particular, we have $b_1 = \cdots = b_r = 0$ as required. □

Remark 12. The above proof never used the fact that W is finite dimensional. Indeed, the statement is still true even when V is finite dimensional but W is not.

Example 13. Recall that the image and kernel of the projection π in Example 2(c) are the x-y-plane and z-axis, respectively. This is consistent with the rank-nullity theorem, because the dimensions of those two subspaces sum to 3, which is the dimension of π's domain $\mathbb{R}^3$.

The rank-nullity theorem can be used to get some remarkably powerful information about T in the case where V and W have the same dimension. As a step in that direction, we now show that the injectivity of T is detected by its kernel.

Lemma 14. *Let V and W be vector spaces and let $T\colon V \to W$ be a linear transformation. Then T is injective if and only if* $\mathrm{Ker}(T) = \{0_V\}$.

Proof. Assume that T is injective and let $v \in \mathrm{Ker}(T)$. We have $T(v) = 0_W$. By Proposition 5, we also have $T(0_V) = 0_W$. Because T is injective, this forces $v = 0_V$, so $\mathrm{Ker}(T) = \{0_V\}$.

Conversely, assume that $\mathrm{Ker}(T) = \{0_V\}$. Let $v, v' \in V$ such that $T(v) = T(v')$. Since T is a linear transformation, we then have $T(v - v') = 0_W$. This means that $v - v' \in \mathrm{Ker}(T)$ so that $v - v' = 0_V$. This shows that $v = v'$. □

The following is the essential ingredient in the invertible matrix theorem.

Theorem 15. *Let* $\dim V = \dim W = n$ *and let* $T : V \to W$ *be a linear transformation. The following statements are equivalent.*

(a) T *is injective.*
(b) T *is surjective.*
(c) T *is bijective.*

Proof. We have the following 'if and only if' statements.

$$\begin{aligned} T \text{ is injective} &\iff \mathrm{Ker}(T) = \{0_V\} \quad \text{(by Lemma 14)} \\ &\iff \mathrm{nullity}(T) = 0 \\ &\iff \mathrm{rank}(T) = n \qquad \text{(by Theorem 11)} \\ &\iff \mathrm{Im}(T) = W \\ &\iff T \text{ is surjective.} \end{aligned}$$

Here the fourth 'if and only if' statement follows by the fact that $\mathrm{Im}(T)$ is a subspace of W, so if it has dimension $n = \dim W$, then it must equal W. □

Linear transformations and matrices

Proposition 16. *Let* V *and* W *be vector spaces with* V *finite dimensional and let* $T : V \to W$ *be a linear transformation. If* $v_1, \dots, v_n$ *is a fixed basis of* V*, then* T *is uniquely determined by the vectors* $T(v_1), \dots, T(v_n) \in W$*.*

Proof. A function is completely determined by where it maps each element of its domain. In order to prove that $T(v_1), \dots, T(v_n)$ determine T, it therefore suffices to show that they allow us to calculate $T(v)$ for each $v \in V$. Observe that since $v_1, \dots, v_n$ is a basis of V, we may write any element $v \in V$ as a linear combination $v = a_1v_1 + \cdots + a_nv_n$. This yields

$$T(v) = T(a_1v_1 + \cdots + a_nv_n) = T(a_1v_1) + \cdots + T(a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n).$$

(The second and third equalities follow from the fact that T is a linear transformation.) It is apparent from this that the only information required to compute $T(v)$ are the values of $T(v_1), \dots, T(v_n)$.

The 'uniquely' part of the proposition has to do with the following: Given any collection of vectors $w_1, \dots, w_n \in W$, there are, in general, many functions $T : V \to W$ that satisfy

$T(v_j) = w_j$ for all j. But there is only one *linear transformation* $T: V \to W$ with $T(v_j) = w_j$ for all j.

To see this, suppose that $T: V \to W$ and $S: V \to W$ are any two linear transformations such that $T(v_j) = S(v_j) = w_j$ for all j. Letting

$$v = a_1 v_1 + \cdots + a_n v_n$$

be any vector in V, we then have

$$T(v) = a_1 T(v_1) + \cdots + a_n T(v_n) = a_1 w_1 + \cdots + a_n w_n = a_1 S(v_1) + \cdots + a_n S(v_n) = S(v).$$

This shows that $T = S$. □

We now consider the case where both V and W are finite dimensional vector spaces. In this case, we can fix bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, respectively. Because the vectors $T(v_j)$ lie in W, we can write them as linear combinations of the basis vectors w_i for W in a unique way. In other words, for each $1 \leq j \leq n$, we have a unique expression

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

This means for each $1 \leq j \leq n$ and each $1 \leq i \leq m$ that a_{ij} is the coefficient on the basis vector w_i in the expression of $T(v_j)$ as a linear combination of $w_1, \ldots, w_m$. The coefficients a_{ij} may be put into a matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Evidently A is an $m \times n$ matrix. We say that A *represents* the linear transformation T with respect to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$. It is very important to note that a different choice of bases for V and W would yield a different matrix representing T in general. In other words, when we talk about a certain matrix representing a linear transformation, we must always assume that bases have already been chosen for V and W.

The converse to the above discussion would be the following: Suppose that we are given an n-dimensional vector space V and an m-dimensional vector space W with chosen bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, respectively. Given any $m \times n$ matrix $A = (a_{ij})$, we may ask whether or not there exists a linear transformation $T: V \to W$ that is represented by A with respect to these chosen bases. The answer is yes.

To construct T, we would first define

$$
\begin{aligned}
T(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m, \\
T(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m, \\
&\vdots \\
T(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m.
\end{aligned} \tag{1}
$$

Now if $v = a_1v_1 + \cdots + a_nv_n$ were any vector in V, we could then define

$$T(v) = a_1T(v_1) + \cdots + a_nT(v_n).$$

It is a routine exercise to verify that the function T defined in this way is actually a linear transformation. It is then clear from the equations (1) that A represents T with respect to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$. All of these observations prove the following result.

Theorem 17. *Let V and W be finite dimensional vector spaces and fix bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, respectively. Then there is one-to-one correspondence*

$$\{\text{linear transformations } V \to W\} \longleftrightarrow \{m \times n \text{ matrices with entries in } F\}.$$

Here, a linear transformation $T\colon V \to W$ is mapped to the $m \times n$ matrix A that represents T with respect to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$.

Now consider the special case of a linear transformation $T\colon F^n \to F^m$. In what follows, we shall work exclusively with the standard bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{f}_1, \ldots, \mathbf{f}_m$ of F^n and F^m, respectively. (Recall that the standard basis of F^r is defined to be the collection of vectors in F^r consisting of the columns of the $r \times r$ identity matrix.) With respect to these bases, there is a more explicit relationship between T and the matrix $A = (a_{ij})$ that represents T. To see this, we first write A in terms of its columns

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix}.$$

For any $1 \leq j \leq n$ we then have

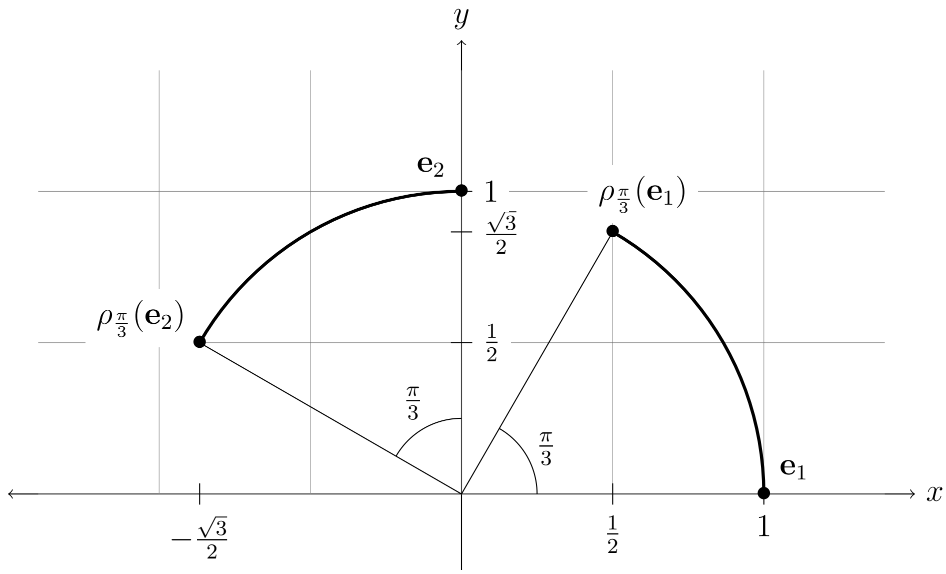
$$T(\mathbf{e}_j) = a_{1j}\mathbf{f}_1 + a_{2j}\mathbf{f}_2 + \cdots + a_{mj}\mathbf{f}_m = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = \mathbf{a}_j.$$

In summary, we have the following proposition.

Proposition 18. *Let $T\colon F^n \to F^m$ be a linear transformation and A the $m \times n$ matrix that represents T with respect to the standard bases of F^n and F^m. Then*

$$A = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{pmatrix}.$$

Example 19. Consider the counter clockwise rotation $\rho_{\frac{\pi}{3}} : \mathbb{R}^2 \to \mathbb{R}^2$ of the x-y-plane about the origin by $\frac{\pi}{3}$ radians. The picture



illustrates that

$$\rho_{\frac{\pi}{3}}(\mathbf{e}_1) = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix} \qquad \text{and} \qquad \rho_{\frac{\pi}{3}}(\mathbf{e}_2) = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}.$$

It follows that the matrix

$$\begin{pmatrix} \rho_{\frac{\pi}{3}}(\mathbf{e}_1) & \rho_{\frac{\pi}{3}}(\mathbf{e}_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$$

represents $\rho_{\frac{\pi}{3}}$ with respect to the standard basis of $\mathbb{R}^2$.

Example 20. Similarly, one may verify that the projection $\pi \colon \mathbb{R}^3 \to \mathbb{R}^3$ defined in Example 2(c) is represented with respect to the standard bases by the 3×3 matrix

$$\begin{pmatrix} \pi(\mathbf{e}_1) & \pi(\mathbf{e}_2) & \pi(\mathbf{e}_3) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Operations on matrices

For the purpose of this section we now fix a bit of notation. If $T \colon V \to W$ is a linear transformation and $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$ are bases of V and W, respectively, we denote the $m \times n$ matrix that represents T with respect to $\mathcal{B}$ and $\mathcal{C}$ by $M_{\mathcal{B}}^{\mathcal{C}}(T)$.

Observe that, up to this point, the only reason we have talked about matrices is because we have been interested in studying linear transformations. Indeed, this is the right attitude to have. Our goal is to motivate certain matrix theoretic operations in terms of common operations on linear transformations.

We first consider addition of linear transformations. Let $T, S\colon V \to W$ be linear transformations. We define a new function $T + S\colon V \to W$ by the rule

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

Note that the right hand side is simply the sum of the two vectors $T(v)$ and $S(v)$ in W. It is straightforward to check that $T + S$ is again a linear transformation called the *sum* of T and S.

Our second operation on linear transformations is scalar multiplication. If $T\colon V \to W$ is a linear transformation and $c \in F$ is a scalar, we define a new function $cT\colon V \to W$ by the rule

$$(cT)(v) = cT(v) \qquad \text{for all } v \in V.$$

Here the right hand side is simply the scalar multiple of the vector $T(v)$ in W by c. Again, it is straightforward to check that cT is a linear transformation, which we call the *scalar multiple* of T by c.

We now let $\mathcal{L}(V, W)$ denote the set of all linear transformations from V to W. One may check that the above defined operations of addition and scalar multiplication give $\mathcal{L}(V, W)$ the natural structure of an F-vector space. It is enlightening to see how that vector space structure translates to the world of matrices.

In this direction, let $A = M_{\mathcal{B}}^{\mathcal{C}}(T)$ and $B = M_{\mathcal{B}}^{\mathcal{C}}(S)$, the $m \times n$ matrices that represent T and S with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. We define $A + B$ to be the $m \times n$ matrix

$$A + B = M_{\mathcal{B}}^{\mathcal{C}}(T + S)$$

that represents $T + S$ with respect to the same bases. This is an abstract definition, but it is not difficult to figure out how to compute $A + B$ concretely from the above information. Write

$$A = (a_{ij}), \qquad B = (b_{ij}), \qquad A + B = (c_{ij}).$$

To compute $A + B$, we need to find the formula for each entry c_{ij}. Since $A + B$ represents $T + S$, we have the relations

$$(T + S)(v_j) = \sum_{i=1}^{m} c_{ij} w_i \qquad \text{for all } 1 \leq j \leq n.$$

On the other hand, because A represents T and B represents S, we have

$$(T + S)(v_j) = T(v_j) + S(v_j) = \left(\sum_{i=1}^{m} a_{ij} w_i \right) + \left(\sum_{i=1}^{m} b_{ij} w_i \right) = \sum_{i=1}^{m} (a_{ij} + b_{ij}) w_i,$$

hence

$$\sum_{i=1}^{m} c_{ij} w_i = \sum_{i=1}^{m} (a_{ij} + b_{ij}) w_i.$$

Since the w_i are linearly independent, this forces $c_{ij} = a_{ij} + b_{ij}$ for all i, j. In summary we obtain the following proposition, which is usually given as the definition of $A + B$.

Proposition 21. *If $A = (a_{ij})$ and $B = (b_{ij})$ are $m \times n$ matrices, then*

$$A + B = (a_{ij} + b_{ij}).$$

Similarly, for a scalar $c \in F$ we define cA to be the $m \times n$ matrix

$$cA = M_{\mathcal{B}}^{\mathcal{C}}(cT)$$

that represents the linear transformation cT. To find the formula for cA we write

$$A = (a_{ij}), \qquad cA = (d_{ij}).$$

Since cA represents cT, we have

$$(cT)(v_j) = \sum_{i=1}^{m} d_{ij} w_i \qquad \text{for all } 1 \leq j \leq n.$$

On the other hand, because A represents T we have

$$(cT)(v_j) = cT(v_j) = c \sum_{i=1}^{m} a_{ij} w_i = \sum_{i=1}^{m} (ca_{ij}) w_i$$

so that

$$\sum_{i=1}^{m} d_{ij} w_i = \sum_{i=1}^{m} (ca_{ij}) w_i$$

and $d_{ij} = ca_{ij}$ for all i, j. In summary we obtain the following proposition, which is usually given as the definition of cA.

Proposition 22. *If $A = (a_{ij})$ is an $m \times n$ matrix, then*

$$cA = (ca_{ij}).$$

There are numerous properties of matrix addition and scalar multiplication that are discussed in most textbooks, but they can all be seen abstractly from the fact that $\mathcal{L}(V, W)$ is a vector space. For example, if one wanted to prove that scalar multiplication of matrices distributes over addition of matrices, one could notice from the above setup that

$$\begin{aligned} c(A + B) &= cM_{\mathcal{B}}^{\mathcal{C}}(T + S) = M_{\mathcal{B}}^{\mathcal{C}}(c(T + S)) = M_{\mathcal{B}}^{\mathcal{C}}(cT + cS) = M_{\mathcal{B}}^{\mathcal{C}}(cT) + M_{\mathcal{B}}^{\mathcal{C}}(cS) \\ &= cA + cB. \end{aligned}$$

All of these equalities are simply the above definitions, except for the third, which follows from the fact that scalar multiplication distributes over addition in $\mathcal{L}(V, W)$.

Our final operation on linear transformations is composition. Recall that if U, V and W are any vector spaces and $S\colon U \to V$ and $T\colon V \to W$ are linear transformations, then the composition $T \circ S\colon U \to W$ is again a linear transformation.

The matrix theoretic operation that corresponds to composition of linear transformations will be matrix multiplication. Let U, V and W be finite dimensional vector spaces and let $S\colon U \to V$ and $T\colon V \to W$ be linear transformations. Choose bases

$$\mathcal{A} = \{u_1, \dots, u_r\}, \qquad \mathcal{B} = \{v_1, \dots, v_n\}, \qquad \mathcal{C} = \{w_1, \dots, w_m\}$$

of U, V and W respectively. Let $B = M_{\mathcal{A}}^{\mathcal{B}}(S)$ be the $n \times r$ matrix that represents S with respect to $\mathcal{A}$ and $\mathcal{B}$, and let $A = M_{\mathcal{B}}^{\mathcal{C}}(T)$ be the $m \times n$ matrix that represents T with respect to $\mathcal{B}$ and $\mathcal{C}$. We define AB to be the matrix

$$AB = M_{\mathcal{A}}^{\mathcal{C}}(T \circ S)$$

that represents $T \circ S$ with respect to $\mathcal{A}$ and $\mathcal{C}$.

To find a concrete formula for the matrix product AB, write

$$A = (a_{ij}), \qquad B = (b_{jk}), \qquad AB = (c_{ik}).$$

Because AB represents $T \circ S$, we have

$$(T \circ S)(u_k) = \sum_{i=1}^{m} c_{ik} w_i \qquad \text{for all } 1 \le k \le r.$$

On the other hand, because A represents T and B represents S we have

$$\begin{aligned}
(T \circ S)(u_k) &= T(S(u_k)) \\
&= T\left(\sum_{j=1}^{n} b_{jk} v_j\right) \\
&= \sum_{j=1}^{n} b_{jk} T(v_j) \qquad\qquad \text{(since } T \text{ is a linear transformation)} \\
&= \sum_{j=1}^{n} b_{jk} \left(\sum_{i=1}^{m} a_{ij} w_i\right) \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} b_{jk}\right) w_i.
\end{aligned}$$

This forces $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$ for all i, k. In summary we obtain the following proposition, which is usually given as the definition of AB.

Proposition 23. *If $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{jk})$ is an $n \times r$ matrix, then AB is the $m \times r$ matrix whose i, k entry is*

$$\sum_{j=1}^{n} a_{ij} b_{jk}.$$

Isomorphism

Definition 24. A linear transformation $T\colon V \to W$ is an *isomorphism* if T is a bijection. We say that V and W are *isomorphic* if there exists an isomorphism $T\colon V \to W$.

The idea behind this definition is that if V and W are isomorphic via $T\colon V \to W$, then V and W have 'the same structure'. Specifically, since T is a bijection, each vector in V corresponds to a unique vector in W. For v and v' in V, write $w = T(v)$ and $w' = T(v')$. Because T is a linear transformation, we have the following correspondences under T.

$$v \longleftrightarrow w$$
$$v' \longleftrightarrow w'$$
$$v + v' \longleftrightarrow w + w'$$
$$av \longleftrightarrow aw$$

The upshot is that any statement about the structure of V will also be true of the structure of W. Indeed, if $\mathscr{P}(v_1, \dots, v_r)$ is a statement involving the vectors $v_1, \dots, v_r$ in V that depends only on addition and scalar multiplication, then it will be true if and only if $\mathscr{P}(T(v_1), \dots, T(v_r))$ is true in W.

In particular, if V is finite dimensional and $v_1, \dots, v_n$ is a basis of V, then because the statement

$$v_1, \dots, v_n \text{ is a basis}$$

is true in V, the statement

$$T(v_1), \dots, T(v_n) \text{ is a basis}$$

will be true in W. This shows that V and W have the same dimension!

There is a better characterisation of isomorphism that allows us to do more in practice.

Proposition 25. *Let $T\colon V \to W$ be a linear transformation. Then T is an isomorphism if and only if there exists a linear transformation $S\colon W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$.*

Proof. If T is an isomorphism, then it is bijection, so it must have an inverse function, that is, there exists a (unique) *function* $S\colon W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. One can show that any function S satisfying this property must also be a linear transformation.

Conversely, if there is a linear transformation S such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$, then in particular S will be an inverse *function* for T, so T will be a bijection. □

We now introduce the concept of invertibility for matrices. We say that an $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix B such that

$$BA = I_n \qquad \text{and} \qquad AB = I_n,$$

where I_n denotes the $n \times n$ *identity matrix*, i.e., the matrix that has the standard basis as its columns. In this case we call B a (*two sided*) *inverse* for A.

The term 'identity matrix' has to do with the fact that if A is any $n \times n$ matrix, then

$$AI_n = A \qquad \text{and} \qquad I_nA = A.$$

In other words, I_n acts as a multiplicative identity.

Proposition 26. *If B and C are inverses for A, then $B = C$.*

Proof. By associativity of matrix multiplication we have

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

□

Having seen that an invertible matrix A has only one inverse, we may now give the child a name, calling it *the* inverse of A and denoting it by A^{-1}.

Theorem 27. *Let V and W be finite dimensional vector spaces, both having dimension n. Let $\mathcal{B}$ be a basis of V and $\mathcal{C}$ a basis of W. Then a linear transformation $T: V \to W$ is an isomorphism if and only if $M_{\mathcal{B}}^{\mathcal{C}}(T)$ is an invertible matrix.*

Proof. Write $A = M_{\mathcal{B}}^{\mathcal{C}}(T)$.

Suppose that T is an isomorphism. By Proposition 25 there is a linear transformation $S: W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. Let $B = M_{\mathcal{C}}^{\mathcal{B}}(S)$. Since $S \circ T = \mathrm{id}_V$, it follows by the material in the previous section that

$$BA = M_{\mathcal{C}}^{\mathcal{B}}(S)M_{\mathcal{B}}^{\mathcal{C}}(T) = M_{\mathcal{B}}^{\mathcal{B}}(S \circ T) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = I_n.$$

(The last equality is left as an exercise.) Thus B is a left inverse for A. Similarly,

$$AB = M_{\mathcal{B}}^{\mathcal{C}}(T)M_{\mathcal{C}}^{\mathcal{B}}(S) = M_{\mathcal{C}}^{\mathcal{C}}(T \circ S) = M_{\mathcal{C}}^{\mathcal{C}}(\mathrm{id}_W) = I_n,$$

so B is also a right inverse for A, hence A is invertible.

Conversely, suppose that A is invertible. Let $S: W \to V$ be the linear transformation represented by A^{-1} with respect to $\mathcal{C}$ and $\mathcal{B}$. We have

$$M_{\mathcal{B}}^{\mathcal{B}}(S \circ T) = M_{\mathcal{C}}^{\mathcal{B}}(S)M_{\mathcal{B}}^{\mathcal{C}}(T) = A^{-1}A = I_n = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V).$$

The one-to-one correspondence of Theorem 17 then reveals that $S \circ T = \mathrm{id}_V$. Similarly,

$$M_{\mathcal{C}}^{\mathcal{C}}(T \circ S) = M_{\mathcal{B}}^{\mathcal{C}}(T)M_{\mathcal{C}}^{\mathcal{B}}(S) = AA^{-1} = I_n = M_{\mathcal{C}}^{\mathcal{C}}(\mathrm{id}_W),$$

so $T \circ S = \mathrm{id}_W$. By Proposition 25, this shows that T is an isomorphism. □

It follows that when someone says 'invertible matrix', you should immediately think 'a matrix that represents an isomorphism'.

The invertible matrix theorem

For the first part of this section we examine a somewhat special situation. Let A be an $m \times n$ matrix. Observe that any vector $\mathbf{v} \in F^n$ may be viewed as an $n \times 1$ matrix. Since A is $m \times n$ and $\mathbf{v}$ is $n \times 1$, we may compute the product $A\mathbf{v}$, which will be $m \times 1$, i.e., an element of F^m. We let $T\colon F^n \to F^m$ be the function given by

$$T(\mathbf{v}) = A\mathbf{v}.$$

The properties of matrix multiplication allow one to convince themselves that T is a linear transformation. It turns out that A and T share a very special relationship.

Proposition 28. *With respect to the standard bases of F^n and F^m, T is represented by A.*

Proof. Write A in terms of its columns

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{pmatrix}.$$

Let B be the $m \times n$ matrix that represents T with respect to the standard bases and write B in terms of its columns

$$B = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{pmatrix}.$$

For each $1 \leq j \leq n$ we have

$$\mathbf{b}_j = T(\mathbf{e}_j)$$

by Proposition 18. On the other hand we have

$$T(\mathbf{e}_j) = A\mathbf{e}_j = 0_F\mathbf{a}_1 + \cdots + 0_F\mathbf{a}_{j-1} + 1_F\mathbf{a}_j + 0_F\mathbf{a}_{j+1} + \cdots + 0_F\mathbf{a}_n = \mathbf{a}_j$$

so that $\mathbf{b}_j = \mathbf{a}_j$. It follows that $B = A$. $\square$

Our next goal is to find out what properties of A determine whether or not T is injective or surjective.

Lemma 29. *Let A be an $m \times n$ matrix. The linear transformation $T\colon F^n \to F^m$ given by $T(\mathbf{v}) = A\mathbf{v}$ is injective if and only if the columns of A are linearly independent.*

Proof. Denote the columns of A by $\mathbf{a}_1, \dots, \mathbf{a}_n$. We have

$$\begin{aligned} T \text{ is injective} &\iff \operatorname{Ker}(T) = \{\mathbf{0}\} \\ &\iff \text{whenever } A\mathbf{v} = T(\mathbf{v}) = \mathbf{0} \text{ then } \mathbf{v} = \mathbf{0} \\ &\iff \text{the matrix equation } A\mathbf{x} = \mathbf{0} \text{ has only the zero solution} \\ &\iff \text{the vector equation } x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \text{ has only the zero solution} \\ &\iff \mathbf{a}_1, \dots, \mathbf{a}_n \text{ are linearly independent.} \end{aligned}$$

$\square$

Lemma 30. *Let A be an $m \times n$ matrix. The linear transformation $T\colon F^n \to F^m$ given by $T(\mathbf{v}) = A\mathbf{v}$ is surjective if and only if the columns of A span F^m.*

Proof. Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$. We have

$$\begin{aligned} T \text{ is surjective} &\iff \text{for each } \mathbf{b} \in F^m \text{ there exists } \mathbf{v} \in F^n \text{ such that } A\mathbf{v} = T(\mathbf{v}) = \mathbf{b} \\ &\iff \text{for each } \mathbf{b} \in F^m \text{ the equation } A\mathbf{x} = \mathbf{b} \text{ has a solution} \\ &\iff \text{for each } \mathbf{b} \in F^m \text{ the equation } x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \text{ has a solution} \\ &\iff \text{each } \mathbf{b} \in F^m \text{ is a linear combination of } \mathbf{a}_1, \ldots, \mathbf{a}_n \\ &\iff \mathbf{a}_1, \ldots, \mathbf{a}_n \text{ span } F^m. \end{aligned}$$

□

We are now in a position to prove the invertible matrix theorem, a highlight of any linear algebra course.

Theorem 31 (The invertible matrix theorem)**.** *Let A be an $n \times n$ matrix and $T\colon F^n \to F^n$ the linear transformation given by $T(\mathbf{v}) = A\mathbf{v}$. The following statements are equivalent.*

(a) *A is invertible.*
(b) *A is row equivalent to I_n.*
(c) *Every echelon form of A has n pivots.*
(d) *The equation $A\mathbf{x} = \mathbf{0}$ has only the zero solution. (A is non-singular.)*
(e) *The columns of A are linearly independent.*
(f) *T is injective.*
(g) *For each $\mathbf{b} \in F^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.*
(h) *The columns of A span F^n.*
(i) *T is surjective.*
(j) *There exists an $n \times n$ matrix C such that $CA = I_n$. (A has a left inverse.)*
(k) *There exists an $n \times n$ matrix D such that $AD = I_n$. (A has a right inverse.)*

Proof. Lemma 29 (and its proof) gives us the equivalences

$$\text{(d)} \iff \text{(e)} \iff \text{(f)},$$

and Lemma 30 gives us

$$\text{(g)} \iff \text{(h)} \iff \text{(i)}.$$

Theorem 15 tells us that (f) and (i) are both equivalent to the condition that T is bijective. Since A represents T with respect to the standard basis of F^n, Theorem 27 tells us that T is bijective if and only if A is invertible. In other words,

$$\text{(f)} \iff \text{(i)} \iff \text{(a)}$$

We now show that (b) and (c) are equivalent to (e). Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ denote the columns of A. We know that these are linearly independent if and only if the vector equation

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

has no free variable. Let E be any echelon form of A. Then $(E\ \mathbf{0})$ is an echelon form of the augmented matrix $(A\ \mathbf{0})$. The above vector equation has no free variable if and only if the columns of E (which correspond to the variables $x_1, \ldots, x_n$) all contain pivots. This establishes the equivalence

$$\text{(e)} \iff \text{(c)}.$$

The equivalence

$$\text{(c)} \iff \text{(b)}$$

follows from the observation that the echelon form E contains n pivot columns if and only if its reduced echelon form (and hence that of A) is I_n.

Finally, we show that (a) is equivalent to (j) and (k). Note that if A is invertible, then (j) and (k) will hold simply by taking $C = D = A^{-1}$. Now assume that A has a left inverse C. Then C represents the linear transformation $L\colon F^n \to F^n$ given by $L(\mathbf{v}) = C\mathbf{v}$ with respect to the standard basis $\mathcal{B}$ of F^n. We have

$$M_{\mathcal{B}}^{\mathcal{B}}(L \circ T) = M_{\mathcal{B}}^{\mathcal{B}}(L)M_{\mathcal{B}}^{\mathcal{B}}(T) = CA = I_n = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_{F^n}).$$

It follows that $L \circ T = \mathrm{id}_{F^n}$. Note that id_{F^n} is injective, and since the composition $L \circ T$ is equal to it, the first map T must also be injective. But we showed that T being injective is equivalent to A being invertible. This establishes the equivalence

$$\text{(a)} \iff \text{(j)}.$$

Similarly, if A has a right inverse D, then since D represents the the linear transformation $R\colon F^n \to F^n$ given by $R(\mathbf{v}) = D\mathbf{v}$ with respect to $\mathcal{B}$, we will have

$$M_{\mathcal{B}}^{\mathcal{B}}(T \circ R) = M_{\mathcal{B}}^{\mathcal{B}}(T)M_{\mathcal{B}}^{\mathcal{B}}(R) = AD = I_n = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_{F^n})$$

so that $T \circ R = \mathrm{id}_{F^n}$. Note that id_{F^n} is surjective, and since the composition $T \circ R$ is equal to it, the second map T must also be surjective. But we showed that T being surjective is equivalent to A being invertible. This establishes the equivalence

$$\text{(a)} \iff \text{(k)}. \qquad \square$$

Change of basis and similar matrices

Let V be a finite dimensional vector space. We now want to examine how linear transformations behave with respect to different bases of V. So let

$$\mathcal{B} = \{v_1, \ldots, v_n\} \qquad \text{and} \qquad \mathcal{C} = \{w_1, \ldots, w_n\}$$

be bases of V. Since $\mathcal{B}$ is a basis, each vector w_j can be written as a linear combination

$$w_j = \sum_{i=1}^{n} p_{ij} v_i.$$

The $n \times n$ matrix (p_{ij}) is called a *change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$*. Since it is an $n \times n$ matrix, we know it represents a liner transformation $V \to V$ with respect to some bases of V. But which transformation, and with respect to which bases?

Proposition 32. *Let V, $\mathcal{B}$ and $\mathcal{C}$ be as above and let $P = (p_{ij})$ be the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$. Then $P = M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_V)$. In particular, P is invertible.*

Proof. The identity on V maps each w_j to itself, so $\mathrm{id}_V(w_j) = w_j = \sum_{i=1}^{n} p_{ij} v_i$. This shows that P represents id_V with respect to the bases $\mathcal{C}$ and $\mathcal{B}$, respectively. $\square$

What is all of this good for? Imagine we have a linear transformation $T\colon V \to V$. We might be given a basis $\mathcal{B}$, so we could compute the matrix $A = M_{\mathcal{B}}^{\mathcal{B}}(T)$ representing T with respect to the bases $\mathcal{B}$. Now suppose we have a friend who prefers to view the world with respect to the bases $\mathcal{C}$. How might we compute how they interpret the information coming from T? In other words, how can we compute the matrix $B = M_{\mathcal{C}}^{\mathcal{C}}(T)$, given that we know A? Here's how.

Note that we have the following diagram of functions.

$$\begin{array}{ccccc}
\mathcal{B} & V & \xrightarrow{T} & V & \mathcal{B} \\
 & \mathrm{id}_V \uparrow & & \uparrow \mathrm{id}_V & \\
\mathcal{C} & V & \xrightarrow[T]{} & V & \mathcal{C}
\end{array}$$

We are interested in computing the matrix representing the bottom arrow with respect to $\mathcal{C}$. We see that this arrow is the linear transformation $T = \mathrm{id}_V^{-1} \circ T \circ \mathrm{id}_V$. The matrix that represents this composition with respect to $\mathcal{C}$ is

$$M_{\mathcal{C}}^{\mathcal{C}}(\mathrm{id}_V^{-1} \circ T \circ \mathrm{id}_V) = M_{\mathcal{B}}^{\mathcal{C}}(\mathrm{id}_V^{-1}) M_{\mathcal{B}}^{\mathcal{B}}(T) M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_V) = M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_V)^{-1} M_{\mathcal{B}}^{\mathcal{B}}(T) M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_V) = P^{-1}AP.$$

Therefore $B = P^{-1}AP$. Matrices that satisfy this sort of relationship have a special name.

Definition 33. Let A and B be $n \times n$ matrices. We say that B *is similar to* A if there is some invertible $n \times n$ matrix P such that $B = P^{-1}AP$.

If an $n \times n$ matrix B is similar to an $n \times n$ matrix A, then B has many things in common with A, mainly because B represents the same linear transformation as A does, except with respect to a different basis. We will see several examples of this later in the course.