Chapter 5

Direct Sums, Rank-Nullity Theorem, Affine Maps

In this chapter all vector spaces are defined over an arbitrary field K. For the sake of concreteness, the reader may safely assume that $K = \mathbb{R}$.

5.1 Direct Products

There are some useful ways of forming new vector spaces from older ones.

Definition 5.1. Given $p \geq 2$ vector spaces $E_1, \ldots, E_p$, the product $F = E_1 \times \cdots \times E_p$ can be made into a vector space by defining addition and scalar multiplication as follows:

$$\begin{aligned}(u_1, \ldots, u_p) + (v_1, \ldots, v_p) &= (u_1 + v_1, \ldots, u_p + v_p) \\ \lambda(u_1, \ldots, u_p) &= (\lambda u_1, \ldots, \lambda u_p),\end{aligned}$$

for all $u_i, v_i \in E_i$ and all $\lambda \in \mathbb{R}$. The zero vector of $E_1 \times \cdots \times E_p$ is the p-tuple

$$(\underbrace{0, \ldots, 0}_{p}),$$

where the ith zero is the zero vector of E_i.

With the above addition and multiplication, the vector space $F = E_1 \times \cdots \times E_p$ is called the *direct product* of the vector spaces $E_1, \ldots, E_p$.

As a special case, when $E_1 = \cdots = E_p = \mathbb{R}$, we find again the vector space $F = \mathbb{R}^p$. The *projection maps* $pr_i \colon E_1 \times \cdots \times E_p \to E_i$ given by

$$pr_i(u_1, \ldots, u_p) = u_i$$

are clearly linear. Similarly, the maps $\mathrm{in}_i \colon E_i \to E_1 \times \cdots \times E_p$ given by

$$\mathrm{in}_i(u_i) = (0, \ldots, 0, u_i, 0, \ldots, 0)$$

are injective and linear. If $\dim(E_i) = n_i$ and if $(e^i_1, \ldots, e^i_{n_i})$ is a basis of E_i for $i = 1, \ldots, p$, then it is easy to see that the $n_1 + \cdots + n_p$ vectors

$$\begin{array}{ccc}
(e^1_1, 0, \ldots, 0), & \ldots, & (e^1_{n_1}, 0, \ldots, 0), \\
\vdots & \vdots & \vdots \\
(0, \ldots, 0, e^i_1, 0, \ldots, 0), & \ldots, & (0, \ldots, 0, e^i_{n_i}, 0, \ldots, 0), \\
\vdots & \vdots & \vdots \\
(0, \ldots, 0, e^p_1), & \ldots, & (0, \ldots, 0, e^p_{n_p})
\end{array}$$

form a basis of $E_1 \times \cdots \times E_p$, and so

$$\dim(E_1 \times \cdots \times E_p) = \dim(E_1) + \cdots + \dim(E_p).$$

5.2 Sums and Direct Sums

Let us now consider a vector space E and p subspaces $U_1, \ldots, U_p$ of E. We have a map

$$a \colon U_1 \times \cdots \times U_p \to E$$

given by

$$a(u_1, \ldots, u_p) = u_1 + \cdots + u_p,$$

with $u_i \in U_i$ for $i = 1, \ldots, p$. It is clear that this map is linear, and so its image is a subspace of E denoted by

$$U_1 + \cdots + U_p$$

and called the *sum* of the subspaces $U_1, \ldots, U_p$. By definition,

$$U_1 + \cdots + U_p = \{u_1 + \cdots + u_p \mid u_i \in U_i,\ 1 \leq i \leq p\},$$

and it is immediately verified that $U_1 + \cdots + U_p$ is the smallest subspace of E containing $U_1, \ldots, U_p$. This also implies that $U_1 + \cdots + U_p$ does not depend on the order of the factors U_i; in particular,

$$U_1 + U_2 = U_2 + U_1.$$

Definition 5.2. For any vector space E and any $p \geq 2$ subspaces $U_1, \ldots, U_p$ of E, if the map $a \colon U_1 \times \cdots \times U_p \to E$ defined above is injective, then the sum $U_1 + \cdots + U_p$ is called a *direct sum* and it is denoted by

$$U_1 \oplus \cdots \oplus U_p.$$

The space E is the *direct sum* of the subspaces U_i if

$$E = U_1 \oplus \cdots \oplus U_p.$$

If the map a is injective, then by Proposition 2.16 we have $\operatorname{Ker} a = \{(\underbrace{0, \ldots, 0}_{p})\}$ where each 0 is the zero vector of E, which means that if $u_i \in U_i$ for $i = 1, \ldots, p$ and if

$$u_1 + \cdots + u_p = 0,$$

then $(u_1, \ldots, u_p) = (0, \ldots, 0)$, that is, $u_1 = 0, \ldots, u_p = 0$.

Proposition 5.1. *If the map $a\colon U_1 \times \cdots \times U_p \to E$ is injective, then every $u \in U_1 + \cdots + U_p$ has a unique expression as a sum*

$$u = u_1 + \cdots + u_p,$$

with $u_i \in U_i$, for $i = 1, \ldots, p$.

Proof. If

$$u = v_1 + \cdots + v_p = w_1 + \cdots + w_p,$$

with $v_i, w_i \in U_i$, for $i = 1, \ldots, p$, then we have

$$w_1 - v_1 + \cdots + w_p - v_p = 0,$$

and since $v_i, w_i \in U_i$ and each U_i is a subspace, $w_i - v_i \in U_i$. The injectivity of a implies that $w_i - v_i = 0$, that is, $w_i = v_i$ for $i = 1, \ldots, p$, which shows the uniqueness of the decomposition of u. $\square$

Proposition 5.2. *If the map $a\colon U_1 \times \cdots \times U_p \to E$ is injective, then any p nonzero vectors $u_1, \ldots, u_p$ with $u_i \in U_i$ are linearly independent.*

Proof. To see this, assume that

$$\lambda_1 u_1 + \cdots + \lambda_p u_p = 0$$

for some $\lambda_i \in \mathbb{R}$. Since $u_i \in U_i$ and U_i is a subspace, $\lambda_i u_i \in U_i$, and the injectivity of a implies that $\lambda_i u_i = 0$, for $i = 1, \ldots, p$. Since $u_i \neq 0$, we must have $\lambda_i = 0$ for $i = 1, \ldots, p$; that is, $u_1, \ldots, u_p$ with $u_i \in U_i$ and $u_i \neq 0$ are linearly independent. $\square$

Observe that if a is injective, then we must have $U_i \cap U_j = (0)$ whenever $i \neq j$. However, this condition is generally not sufficient if $p \geq 3$. For example, if $E = \mathbb{R}^2$ and U_1 the line spanned by $e_1 = (1, 0)$, U_2 is the line spanned by $d = (1, 1)$, and U_3 is the line spanned by $e_2 = (0, 1)$, then $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{(0, 0)\}$, but $U_1 + U_2 = U_1 + U_3 = U_2 + U_3 = \mathbb{R}^2$, so $U_1 + U_2 + U_3$ is not a direct sum. For example, d is expressed in two different ways as

$$d = (1, 1) = (1, 0) + (0, 1) = e_1 + e_2.$$

See Figure 5.1.

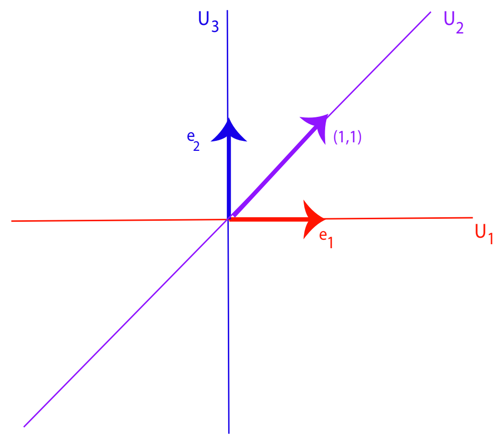

Figure 5.1: The linear subspaces U_1, U_2, and U_3 illustrated as lines in $\mathbb{R}^2$.

As in the case of a sum, $U_1 \oplus U_2 = U_2 \oplus U_1$. *Observe that when the map a is injective, then it is a linear isomorphism between $U_1 \times \cdots \times U_p$ and $U_1 \oplus \cdots \oplus U_p$.* The difference is that $U_1 \times \cdots \times U_p$ is defined even if the spaces U_i are not assumed to be subspaces of some common space.

If E is a direct sum $E = U_1 \oplus \cdots \oplus U_p$, since any p nonzero vectors $u_1, \ldots, u_p$ with $u_i \in U_i$ are linearly independent, if we pick a basis $(u_k)_{k \in I_j}$ in U_j for $j = 1, \ldots, p$, then $(u_i)_{i \in I}$ with $I = I_1 \cup \cdots \cup I_p$ is a basis of E. Intuitively, E is split into p independent subspaces.

Conversely, given a basis $(u_i)_{i \in I}$ of E, if we partition the index set I as $I = I_1 \cup \cdots \cup I_p$, then each subfamily $(u_k)_{k \in I_j}$ spans some subspace U_j of E, and it is immediately verified that we have a direct sum

$$E = U_1 \oplus \cdots \oplus U_p.$$

Definition 5.3. Let $f\colon E \to E$ be a linear map. For any subspace U of E, if $f(U) \subseteq U$ we say that U is *invariant under f*.

Assume that E is finite-dimensional, a direct sum $E = U_1 \oplus \cdots \oplus U_p$, and that each U_j is invariant under f. If we pick a basis $(u_i)_{i \in I}$ as above with $I = I_1 \cup \cdots \cup I_p$ and with each $(u_k)_{k \in I_j}$ a basis of U_j, since each U_j is invariant under f, the image $f(u_k)$ of every basis vector u_k with $k \in I_j$ belongs to U_j, so the matrix A representing f over the basis $(u_i)_{i \in I}$ is a *block diagonal* matrix of the form

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p \end{pmatrix},$$

with each block A_j a $d_j \times d_j$-matrix with $d_j = \dim(U_j)$ and all other entries equal to 0. If $d_j = 1$ for $j = 1, \ldots, p$, the matrix A is a diagonal matrix.

There are natural injections from each U_i to E denoted by $\mathrm{in}_i \colon U_i \to E$.

Now, if $p = 2$, it is easy to determine the kernel of the map $a \colon U_1 \times U_2 \to E$. We have

$$a(u_1, u_2) = u_1 + u_2 = 0 \quad \text{iff} \quad u_1 = -u_2,\ u_1 \in U_1, u_2 \in U_2,$$

which implies that

$$\mathrm{Ker}\, a = \{(u, -u) \mid u \in U_1 \cap U_2\}.$$

Now, $U_1 \cap U_2$ is a subspace of E and the linear map $u \mapsto (u, -u)$ is clearly an isomorphism between $U_1 \cap U_2$ and $\mathrm{Ker}\, a$, so $\mathrm{Ker}\, a$ is isomorphic to $U_1 \cap U_2$. As a consequence, we get the following result:

Proposition 5.3. *Given any vector space E and any two subspaces U_1 and U_2, the sum $U_1 + U_2$ is a direct sum iff $U_1 \cap U_2 = (0)$.*

An interesting illustration of the notion of direct sum is the decomposition of a square matrix into its symmetric part and its skew-symmetric part. Recall that an $n \times n$ matrix $A \in \mathrm{M}_n$ is *symmetric* if $A^\top = A$, *skew -symmetric* if $A^\top = -A$. It is clear that s

$$\mathbf{S}(n) = \{A \in \mathrm{M}_n \mid A^\top = A\} \quad \text{and} \quad \mathbf{Skew}(n) = \{A \in \mathrm{M}_n \mid A^\top = -A\}$$

are subspaces of M_n, and that $\mathbf{S}(n) \cap \mathbf{Skew}(n) = (0)$. Observe that for any matrix $A \in \mathrm{M}_n$, the matrix $H(A) = (A + A^\top)/2$ is symmetric and the matrix $S(A) = (A - A^\top)/2$ is skew-symmetric. Since

$$A = H(A) + S(A) = \frac{A + A^\top}{2} + \frac{A - A^\top}{2},$$

we see that $\mathrm{M}_n = \mathbf{S}(n) + \mathbf{Skew}(n)$, and since $\mathbf{S}(n) \cap \mathbf{Skew}(n) = (0)$, we have the direct sum

$$\mathrm{M}_n = \mathbf{S}(n) \oplus \mathbf{Skew}(n).$$

Remark: The vector space $\mathbf{Skew}(n)$ of skew-symmetric matrices is also denoted by $\mathfrak{so}(n)$. It is the *Lie algebra* of the group $\mathbf{SO}(n)$.

Proposition 5.3 can be generalized to any $p \geq 2$ subspaces at the expense of notation. The proof of the following proposition is left as an exercise.

Proposition 5.4. *Given any vector space E and any $p \geq 2$ subspaces $U_1, \ldots, U_p$, the following properties are equivalent:*

(1) The sum $U_1 + \cdots + U_p$ is a direct sum.

(2) We have

$$U_i \cap \left(\sum_{j=1, j \neq i}^{p} U_j \right) = (0), \quad i = 1, \ldots, p.$$

(3) We have

$$U_i \cap \left(\sum_{j=1}^{i-1} U_j \right) = (0), \quad i = 2, \ldots, p.$$

Because of the isomorphism

$$U_1 \times \cdots \times U_p \approx U_1 \oplus \cdots \oplus U_p,$$

we have

Proposition 5.5. *If E is any vector space, for any (finite-dimensional) subspaces $U_1, \ldots,$ U_p of E, we have*

$$\dim(U_1 \oplus \cdots \oplus U_p) = \dim(U_1) + \cdots + \dim(U_p).$$

If E is a direct sum

$$E = U_1 \oplus \cdots \oplus U_p,$$

since every $u \in E$ can be written in a unique way as

$$u = u_1 + \cdots + u_p$$

with $u_i \in U_i$ for $i = 1 \ldots, p$, we can define the maps $\pi_i \colon E \to U_i$, called *projections*, by

$$\pi_i(u) = \pi_i(u_1 + \cdots + u_p) = u_i.$$

It is easy to check that these maps are linear and satisfy the following properties:

$$\pi_j \circ \pi_i = \begin{cases} \pi_i & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$
$$\pi_1 + \cdots + \pi_p = \mathrm{id}_E.$$

For example, in the case of the direct sum

$$\mathrm{M}_n = \mathbf{S}(n) \oplus \mathbf{Skew}(n),$$

the projection onto $\mathbf{S}(n)$ is given by

$$\pi_1(A) = H(A) = \frac{A + A^\top}{2},$$

and the projection onto $\mathbf{Skew}(n)$ is given by

$$\pi_2(A) = S(A) = \frac{A - A^\top}{2}.$$

Clearly, $H(A)+S(A) = A$, $H(H(A)) = H(A)$, $S(S(A)) = S(A)$, and $H(S(A)) = S(H(A)) = 0$.

A function f such that $f \circ f = f$ is said to be *idempotent*. Thus, the projections π_i are idempotent. Conversely, the following proposition can be shown:

Proposition 5.6. *Let E be a vector space. For any $p \geq 2$ linear maps $f_i : E \to E$, if*

$$f_j \circ f_i = \begin{cases} f_i & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$
$$f_1 + \cdots + f_p = \mathrm{id}_E,$$

then if we let $U_i = f_i(E)$, we have a direct sum

$$E = U_1 \oplus \cdots \oplus U_p.$$

We also have the following proposition characterizing idempotent linear maps whose proof is also left as an exercise.

Proposition 5.7. *For every vector space E, if $f : E \to E$ is an idempotent linear map, i.e., $f \circ f = f$, then we have a direct sum*

$$E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, f,$$

so that f is the projection onto its image $\mathrm{Im}\, f$.

We are now ready to prove a very crucial result relating the rank and the dimension of the kernel of a linear map.

5.3 The Rank-Nullity Theorem; Grassmann's Relation

We begin with the following theorem which shows that given a linear map $f : E \to F$, its domain E is the direct sum of its kernel $\mathrm{Ker}\, f$ with some isomorphic copy of its image $\mathrm{Im}\, f$.

Theorem 5.8. *(Rank-nullity theorem) Let $f : E \to F$ be a linear map with finite image. For any choice of a basis $(f_1, \ldots, f_r)$ of $\mathrm{Im}\, f$, let $(u_1, \ldots, u_r)$ be any vectors in E such that $f_i = f(u_i)$, for $i = 1, \ldots, r$. If $s : \mathrm{Im}\, f \to E$ is the unique linear map defined by $s(f_i) = u_i$, for $i = 1, \ldots, r$, then s is injective, $f \circ s = \mathrm{id}$, and we have a direct sum*

$$E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, s$$

as illustrated by the following diagram:

$$\mathrm{Ker}\, f \longrightarrow E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, s \underset{s}{\overset{f}{\rightleftarrows}} \mathrm{Im}\, f \subseteq F.$$

See Figure 5.2. As a consequence, if E is finite-dimensional, then

$$\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(\mathrm{Ker}\, f) + \mathrm{rk}(f).$$

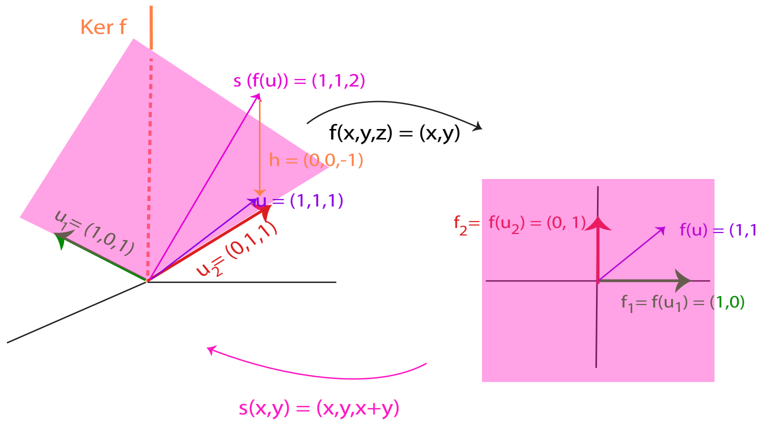

Figure 5.2: Let $f\colon E \to F$ be the linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$ given by $f(x, y, z) = (x, y)$. Then $s\colon \mathbb{R}^2 \to \mathbb{R}^3$ is given by $s(x, y) = (x, y, x + y)$ and maps the pink $\mathbb{R}^2$ isomorphically onto the slanted pink plane of $\mathbb{R}^3$ whose equation is $-x - y + z = 0$. Theorem 5.8 shows that $\mathbb{R}^3$ is the direct sum of the plane $-x - y + z = 0$ and the kernel of f which the orange z-axis.

Proof. The vectors $u_1, \ldots, u_r$ must be linearly independent since otherwise we would have a nontrivial linear dependence

$$\lambda_1 u_1 + \cdots + \lambda_r u_r = 0,$$

and by applying f, we would get the nontrivial linear dependence

$$0 = \lambda_1 f(u_1) + \cdots + \lambda_r f(u_r) = \lambda_1 f_1 + \cdots + \lambda_r f_r,$$

contradicting the fact that $(f_1, \ldots, f_r)$ is a basis. Therefore, the unique linear map s given by $s(f_i) = u_i$, for $i = 1, \ldots, r$, is a linear isomorphism between $\mathrm{Im}\, f$ and its image, the subspace spanned by $(u_1, \ldots, u_r)$. It is also clear by definition that $f \circ s = \mathrm{id}$. For any $u \in E$, let

$$h = u - (s \circ f)(u).$$

Since $f \circ s = \mathrm{id}$, we have

$$\begin{aligned} f(h) = f(u - (s \circ f)(u)) = f(u) - (f \circ s \circ f)(u) & \\ = f(u) - (\mathrm{id} \circ f)(u) = f(u) - f(u) = 0, \end{aligned}$$

which shows that $h \in \mathrm{Ker}\, f$. Since $h = u - (s \circ f)(u)$, it follows that

$$u = h + s(f(u)),$$

with $h \in \mathrm{Ker}\, f$ and $s(f(u)) \in \mathrm{Im}\, s$, which proves that

$$E = \mathrm{Ker}\, f + \mathrm{Im}\, s.$$

Now if $u \in \operatorname{Ker} f \cap \operatorname{Im} s$, then $u = s(v)$ for some $v \in F$ and $f(u) = 0$ since $u \in \operatorname{Ker} f$. Since $u = s(v)$ and $f \circ s = \mathrm{id}$, we get

$$0 = f(u) = f(s(v)) = v,$$

and so $u = s(v) = s(0) = 0$. Thus, $\operatorname{Ker} f \cap \operatorname{Im} s = (0)$, which proves that we have a direct sum

$$E = \operatorname{Ker} f \oplus \operatorname{Im} s.$$

The equation

$$\dim(E) = \dim(\operatorname{Ker} f) + \dim(\operatorname{Im} f) = \dim(\operatorname{Ker} f) + \operatorname{rk}(f)$$

is an immediate consequence of the fact that the dimension is an additive property for direct sums, that by definition the rank of f is the dimension of the image of f, and that $\dim(\operatorname{Im} s) = \dim(\operatorname{Im} f)$, because s is an isomorphism between $\operatorname{Im} f$ and $\operatorname{Im} s$. ☐

Remark: The statement $E = \operatorname{Ker} f \oplus \operatorname{Im} s$ holds if E has infinite dimension. It still holds if $\operatorname{Im}(f)$ also has infinite dimension.

Definition 5.4. The dimension $\dim(\operatorname{Ker} f)$ of the kernel of a linear map f is called the *nullity* of f.

We now derive some important results using Theorem 5.8.

Proposition 5.9. *Given a vector space E, if U and V are any two finite-dimensional subspaces of E, then*

$$\dim(U) + \dim(V) = \dim(U + V) + \dim(U \cap V),$$

an equation known as Grassmann's relation.

Proof. Recall that $U + V$ is the image of the linear map

$$a\colon U \times V \to E$$

given by

$$a(u, v) = u + v,$$

and that we proved earlier that the kernel $\operatorname{Ker} a$ of a is isomorphic to $U \cap V$. By Theorem 5.8,

$$\dim(U \times V) = \dim(\operatorname{Ker} a) + \dim(\operatorname{Im} a),$$

but $\dim(U \times V) = \dim(U) + \dim(V)$, $\dim(\operatorname{Ker} a) = \dim(U \cap V)$, and $\operatorname{Im} a = U + V$, so the Grassmann relation holds. ☐

The Grassmann relation can be very useful to figure out whether two subspace have a nontrivial intersection in spaces of dimension > 3. For example, it is easy to see that in $\mathbb{R}^5$, there are subspaces U and V with $\dim(U) = 3$ and $\dim(V) = 2$ such that $U \cap V = (0)$; for example, let U be generated by the vectors $(1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0)$, and V be generated by the vectors $(0, 0, 0, 1, 0)$ and $(0, 0, 0, 0, 1)$. However, we claim that if $\dim(U) = 3$ and $\dim(V) = 3$, then $\dim(U \cap V) \geq 1$. Indeed, by the Grassmann relation, we have

$$\dim(U) + \dim(V) = \dim(U + V) + \dim(U \cap V),$$

namely

$$3 + 3 = 6 = \dim(U + V) + \dim(U \cap V),$$

and since $U + V$ is a subspace of $\mathbb{R}^5$, $\dim(U + V) \leq 5$, which implies

$$6 \leq 5 + \dim(U \cap V),$$

that is $1 \leq \dim(U \cap V)$.

As another consequence of Proposition 5.9, if U and V are two hyperplanes in a vector space of dimension n, so that $\dim(U) = n - 1$ and $\dim(V) = n - 1$, the reader should show that

$$\dim(U \cap V) \geq n - 2,$$

and so, if $U \neq V$, then

$$\dim(U \cap V) = n - 2.$$

Here is a characterization of direct sums that follows directly from Theorem 5.8.

Proposition 5.10. *If $U_1, \ldots, U_p$ are any subspaces of a finite dimensional vector space E, then*

$$\dim(U_1 + \cdots + U_p) \leq \dim(U_1) + \cdots + \dim(U_p),$$

and

$$\dim(U_1 + \cdots + U_p) = \dim(U_1) + \cdots + \dim(U_p)$$

iff the U_is form a direct sum $U_1 \oplus \cdots \oplus U_p$.

Proof. If we apply Theorem 5.8 to the linear map

$$a \colon U_1 \times \cdots \times U_p \to U_1 + \cdots + U_p$$

given by $a(u_1, \ldots, u_p) = u_1 + \cdots + u_p$, we get

$$\begin{aligned}\dim(U_1 + \cdots + U_p) &= \dim(U_1 \times \cdots \times U_p) - \dim(\operatorname{Ker} a)\\ &= \dim(U_1) + \cdots + \dim(U_p) - \dim(\operatorname{Ker} a),\end{aligned}$$

so the inequality follows. Since a is injective iff $\operatorname{Ker} a = (0)$, the U_is form a direct sum iff the second equation holds. $\square$

Another important corollary of Theorem 5.8 is the following result:

Proposition 5.11. *Let E and F be two vector spaces with the same finite dimension* $\dim(E) = \dim(F) = n$. *For every linear map $f\colon E \to F$, the following properties are equivalent:*

(a) *f is bijective.*

(b) *f is surjective.*

(c) *f is injective.*

(d) $\operatorname{Ker} f = (0)$.

Proof. Obviously, (a) implies (b).

If f is surjective, then $\operatorname{Im} f = F$, and so $\dim(\operatorname{Im} f) = n$. By Theorem 5.8,

$$\dim(E) = \dim(\operatorname{Ker} f) + \dim(\operatorname{Im} f),$$

and since $\dim(E) = n$ and $\dim(\operatorname{Im} f) = n$, we get $\dim(\operatorname{Ker} f) = 0$, which means that $\operatorname{Ker} f = (0)$, and so f is injective (see Proposition 2.16). This proves that (b) implies (c).

If f is injective, then by Proposition 2.16, $\operatorname{Ker} f = (0)$, so (c) implies (d).

Finally, assume that $\operatorname{Ker} f = (0)$, so that $\dim(\operatorname{Ker} f) = 0$ and f is injective (by Proposition 2.16). By Theorem 5.8,

$$\dim(E) = \dim(\operatorname{Ker} f) + \dim(\operatorname{Im} f),$$

and since $\dim(\operatorname{Ker} f) = 0$, we get

$$\dim(\operatorname{Im} f) = \dim(E) = \dim(F),$$

which proves that f is also surjective, and thus bijective. This proves that (d) implies (a) and concludes the proof. $\square$

One should be warned that Proposition 5.11 fails in infinite dimension. A linear map may be injective without being surjective and vice versa.

Here are a few applications of Proposition 5.11. Let A be an $n \times n$ matrix and assume that A some right inverse B, which means that B is an $n \times n$ matrix such that

$$AB = I.$$

The linear map associated with A is surjective, since for every $u \in \mathbb{R}^n$, we have $A(Bu) = u$. By Proposition 5.11, this map is bijective so B is actually the inverse of A; in particular $BA = I$.

Similarly, assume that A has a left inverse B, so that

$$BA = I.$$

This time the linear map associated with A is injective, because if $Au = 0$, then $BAu = B0 = 0$, and since $BA = I$ we get $u = 0$. Again, by Proposition 5.11, this map is bijective so B is actually the inverse of A; in particular $AB = I$.

Now assume that the linear system $Ax = b$ has some solution for every b. Then the linear map associated with A is surjective and by Proposition 5.11, A is invertible.

Finally assume that the linear system $Ax = b$ has at most one solution for every b. Then the linear map associated with A is injective and by Proposition 5.11, A is invertible.

We also have the following basic proposition about injective or surjective linear maps.

Proposition 5.12. *Let E and F be vector spaces, and let $f\colon E \to F$ be a linear map. If $f\colon E \to F$ is injective, then there is a surjective linear map $r\colon F \to E$ called a retraction, such that $r \circ f = \mathrm{id}_E$. See Figure 5.3. If $f\colon E \to F$ is surjective, then there is an injective linear map $s\colon F \to E$ called a section, such that $f \circ s = \mathrm{id}_F$. See Figure 5.2.*

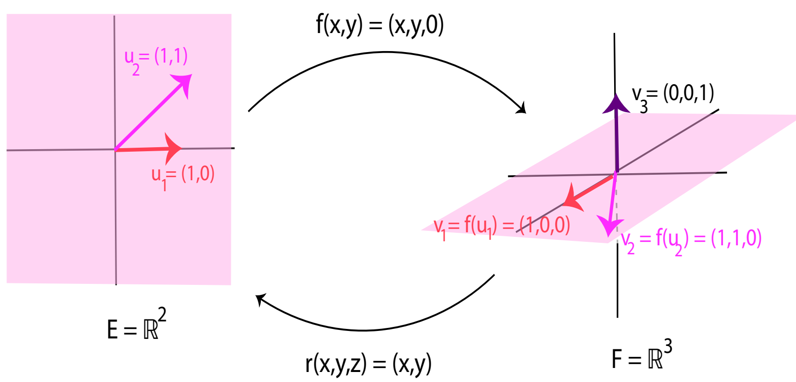

Figure 5.3: Let $f\colon E \to F$ be the injective linear map from $\mathbb{R}^2$ to $\mathbb{R}^3$ given by $f(x, y) = (x, y, 0)$. Then a surjective retraction is given by $r\colon \mathbb{R}^3 \to \mathbb{R}^2$ is given by $r(x, y, z) = (x, y)$. Observe that $r(v_1) = u_1$, $r(v_2) = u_2$, and $r(v_3) = 0$.

Proof. Let $(u_i)_{i\in I}$ be a basis of E. Since $f\colon E \to F$ is an injective linear map, by Proposition 2.17, $(f(u_i))_{i\in I}$ is linearly independent in F. By Theorem 2.9, there is a basis $(v_j)_{j\in J}$ of F, where $I \subseteq J$, and where $v_i = f(u_i)$, for all $i \in I$. By Proposition 2.17, a linear map $r\colon F \to E$ can be defined such that $r(v_i) = u_i$, for all $i \in I$, and $r(v_j) = w$ for all $j \in (J - I)$, where w is any given vector in E, say $w = 0$. Since $r(f(u_i)) = u_i$ for all $i \in I$, by Proposition 2.17, we have $r \circ f = \mathrm{id}_E$.

Now assume that $f\colon E \to F$ is surjective. Let $(v_j)_{j\in J}$ be a basis of F. Since $f\colon E \to F$ is surjective, for every $v_j \in F$, there is some $u_j \in E$ such that $f(u_j) = v_j$. Since $(v_j)_{j\in J}$ is a basis of F, by Proposition 2.17, there is a unique linear map $s\colon F \to E$ such that $s(v_j) = u_j$. Also since $f(s(v_j)) = v_j$, by Proposition 2.17 (again), we must have $f \circ s = \mathrm{id}_F$. $\square$

Remark: Proposition 5.12 also holds if E or F has infinite dimension.

The converse of Proposition 5.12 is obvious.

The notion of rank of a linear map or of a matrix important, both theoretically and practically, since it is the key to the solvability of linear equations. We have the following simple proposition.

Proposition 5.13. *Given a linear map $f\colon E \to F$, the following properties hold:*

(i) $\mathrm{rk}(f) + \dim(\mathrm{Ker}\, f) = \dim(E)$.

(ii) $\mathrm{rk}(f) \leq \min(\dim(E), \dim(F))$.

Proof. Property (i) follows from Proposition 5.8. As for (ii), since $\mathrm{Im}\, f$ is a subspace of F, we have $\mathrm{rk}(f) \leq \dim(F)$, and since $\mathrm{rk}(f) + \dim(\mathrm{Ker}\, f) = \dim(E)$, we have $\mathrm{rk}(f) \leq \dim(E)$. $\square$

The rank of a matrix is defined as follows.

Definition 5.5. Given a $m \times n$-matrix $A = (a_{i\,j})$, the *rank* $\mathrm{rk}(A)$ of the matrix A is the maximum number of linearly independent columns of A (viewed as vectors in $\mathbb{R}^m$).

In view of Proposition 2.10, the rank of a matrix A is the dimension of the subspace of $\mathbb{R}^m$ generated by the columns of A. Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis of E, and $(v_1, \ldots, v_m)$ a basis of F. Let $f\colon E \to F$ be a linear map, and let $M(f)$ be its matrix w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$. Since the rank $\mathrm{rk}(f)$ of f is the dimension of $\mathrm{Im}\, f$, which is generated by $(f(u_1), \ldots, f(u_n))$, the rank of f is the maximum number of linearly independent vectors in $(f(u_1), \ldots, f(u_n))$, which is equal to the number of linearly independent columns of $M(f)$, since F and $\mathbb{R}^m$ are isomorphic. *Thus, we have* $\mathrm{rk}(f) = \mathrm{rk}(M(f))$*, for every matrix representing* f.

We will see later, using duality, that the rank of a matrix A is also equal to the maximal number of linearly independent rows of A.

5.4 Affine Maps

We showed in Section 2.7 that every linear map f must send the zero vector to the zero vector; that is,

$$f(0) = 0.$$

Yet for any fixed nonzero vector $u \in E$ (where E is any vector space), the function t_u given by

$$t_u(x) = x + u, \quad \text{for all} \quad x \in E$$

shows up in practice (for example, in robotics). Functions of this type are called *translations*. They are *not* linear for $u \neq 0$, since $t_u(0) = 0 + u = u$.

More generally, functions combining linear maps and translations occur naturally in many applications (robotics, computer vision, *etc.*), so it is necessary to understand some basic properties of these functions. For this, the notion of affine combination turns out to play a key role.

Recall from Section 2.7 that for any vector space E, given any family $(u_i)_{i\in I}$ of vectors $u_i \in E$, an *affine combination* of the family $(u_i)_{i\in I}$ is an expression of the form

$$\sum_{i\in I} \lambda_i u_i \quad \text{with} \quad \sum_{i\in I} \lambda_i = 1,$$

where $(\lambda_i)_{i\in I}$ is a family of scalars.

A linear combination places no restriction on the scalars involved, but an affine combination is a linear combination *with the restriction that the scalars λ_i must add up to* 1. Nevertheless, a linear combination can always be viewed as an affine combination using the following trick involving 0. For any family $(u_i)_{i\in I}$ of vectors in E and for *any* family of scalars $(\lambda_i)_{i\in I}$, we can write the linear combination $\sum_{i\in I} \lambda_i u_i$ as an affine combination as follows:

$$\sum_{i\in I} \lambda_i u_i = \sum_{i\in I} \lambda_i u_i + \left(1 - \sum_{i\in I} \lambda_i\right) 0.$$

Affine combinations are also called *barycentric combinations*.

Although this is not obvious at first glance, the condition that the scalars λ_i add up to 1 ensures that affine combinations are preserved under translations. To make this precise, consider functions $f\colon E \to F$, where E and F are two vector spaces, such that there is some *linear map* $h\colon E \to F$ and some fixed vector $b \in F$ (a *translation vector*), such that

$$f(x) = h(x) + b, \quad \text{for all} \quad x \in E.$$

The map f given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 8/5 & -6/5 \\ 3/10 & 2/5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is an example of the composition of a linear map with a translation.

We claim that functions of this type preserve affine combinations.

Proposition 5.14. *For any two vector spaces E and F, given any function $f\colon E \to F$ defined such that*

$$f(x) = h(x) + b, \quad \text{for all} \quad x \in E,$$

where $h\colon E \to F$ is a linear map and b is some fixed vector in F, for every affine combination $\sum_{i\in I} \lambda_i u_i$ (with $\sum_{i\in I} \lambda_i = 1$), we have

$$f\left(\sum_{i\in I} \lambda_i u_i\right) = \sum_{i\in I} \lambda_i f(u_i).$$

In other words, f preserves affine combinations.

Proof. By definition of f, using the fact that h is linear and the fact that $\sum_{i\in I}\lambda_i = 1$, we have

$$\begin{aligned} f\biggl(\sum_{i\in}\lambda_i u_i\biggr) &= h\biggl(\sum_{i\in I}\lambda_i u_i\biggr) + b \\ &= \sum_{i\in I}\lambda_i h(u_i) + 1b \\ &= \sum_{i\in I}\lambda_i h(u_i) + \biggl(\sum_{i\in I}\lambda_i\biggr) b \\ &= \sum_{i\in I}\lambda_i (h(u_i) + b) \\ &= \sum_{i\in I}\lambda_i f(u_i), \end{aligned}$$

as claimed. ∎

Observe how the fact that $\sum_{i\in I}\lambda_i = 1$ was used in a crucial way in Line 3. Surprisingly, the converse of Proposition 5.14 also holds.

Proposition 5.15. *For any two vector spaces E and F, let $f\colon E \to F$ be any function that preserves affine combinations, i.e., for every affine combination $\sum_{i\in I}\lambda_i u_i$ (with $\sum_{i\in I}\lambda_i = 1$), we have*

$$f\biggl(\sum_{i\in I}\lambda_i u_i\biggr) = \sum_{i\in I}\lambda_i f(u_i).$$

Then for any $a \in E$, the function $h\colon E \to F$ given by

$$h(x) = f(a + x) - f(a)$$

is a linear map independent of a, and

$$f(a + x) = h(x) + f(a), \quad \text{for all} \quad x \in E.$$

In particular, for $a = 0$, if we let $c = f(0)$, then

$$f(x) = h(x) + c, \quad \text{for all} \quad x \in E.$$

Proof. First, let us check that h is linear. Since f preserves affine combinations and since $a + u + v = (a + u) + (a + v) - a$ is an affine combination $(1 + 1 - 1 = 1)$, we have

$$\begin{aligned} h(u + v) &= f(a + u + v) - f(a) \\ &= f((a + u) + (a + v) - a) - f(a) \\ &= f(a + u) + f(a + v) - f(a) - f(a) \\ &= f(a + u) - f(a) + f(a + v) - f(a) \\ &= h(u) + h(v). \end{aligned}$$

This proves that

$$h(u+v) = h(u) + h(v), \quad u, v \in E.$$

Observe that $a + \lambda u = \lambda(a+u) + (1-\lambda)a$ is also an affine combination ($\lambda + 1 - \lambda = 1$), so we have

$$\begin{aligned} h(\lambda u) &= f(a + \lambda u) - f(a) \\ &= f(\lambda(a+u) + (1-\lambda)a) - f(a) \\ &= \lambda f(a+u) + (1-\lambda) f(a) - f(a) \\ &= \lambda(f(a+u) - f(a)) \\ &= \lambda h(u). \end{aligned}$$

This proves that

$$h(\lambda u) = \lambda h(u), \quad u \in E, \ \lambda \in \mathbb{R}.$$

Therefore, h is indeed linear.

For any $b \in E$, since $b + u = (a+u) - a + b$ is an affine combination ($1 - 1 + 1 = 1$), we have

$$\begin{aligned} f(b+u) - f(b) &= f((a+u) - a + b) - f(b) \\ &= f(a+u) - f(a) + f(b) - f(b) \\ &= f(a+u) - f(a), \end{aligned}$$

which proves that for all $a, b \in E$,

$$f(b+u) - f(b) = f(a+u) - f(a), \quad u \in E.$$

Therefore $h(x) = f(a+u) - f(a)$ does not depend on a, and it is obvious by the definition of h that

$$f(a+x) = h(x) + f(a), \quad \text{for all} \quad x \in E.$$

For $a = 0$, we obtain the last part of our proposition. $\square$

We should think of a as a *chosen origin* in E. The function f maps the origin a in E to the origin $f(a)$ in F. Proposition 5.15 shows that the definition of h does not depend on the origin chosen in E. Also, since

$$f(x) = h(x) + c, \quad \text{for all} \quad x \in E$$

for some fixed vector $c \in F$, we see that f is the composition of the linear map h with the translation t_c (in F).

The unique linear map h as above is called the *linear map associated with* f, and it is sometimes denoted by $\overrightarrow{f}$.

In view of Propositions 5.14 and 5.15, it is natural to make the following definition.

Definition 5.6. For any two vector spaces E and F, a function $f: E \to F$ is an *affine map* if f preserves affine combinations, *i.e.*, for every affine combination $\sum_{i\in I} \lambda_i u_i$ (with $\sum_{i\in I} \lambda_i = 1$), we have

$$f\left(\sum_{i\in I} \lambda_i u_i\right) = \sum_{i\in I} \lambda_i f(u_i).$$

Equivalently, a function $f: E \to F$ is an *affine map* if there is some linear map $h: E \to F$ (also denoted by $\overrightarrow{f}$) and some fixed vector $c \in F$ such that

$$f(x) = h(x) + c, \quad \text{for all} \quad x \in E.$$

Note that a linear map always maps the standard origin 0 in E to the standard origin 0 in F. However an affine map usually maps 0 to a nonzero vector $c = f(0)$. This is the "translation component" of the affine map.

When we deal with affine maps, it is often fruitful to think of the elements of E and F not only as vectors but also as *points*. In this point of view, *points can only be combined using affine combinations*, but vectors can be combined in an unrestricted fashion using linear combinations. We can also think of $u + v$ as the *result of translating the point u by the translation t_v*. These ideas lead to the definition of *affine spaces*.

The idea is that instead of a single space E, an affine space consists of two sets E and $\overrightarrow{E}$, where E is just an unstructured set of points, and $\overrightarrow{E}$ is a vector space. Furthermore, the vector space $\overrightarrow{E}$ acts on E. We can think of $\overrightarrow{E}$ as a set of *translations* specified by vectors, and given any point $a \in E$ and any vector (translation) $u \in \overrightarrow{E}$, the result of translating a by u is the point (not vector) $a + u$. Formally, we have the following definition.

Definition 5.7. An *affine space* is either the degenerate space reduced to the empty set, or a triple $\langle E, \overrightarrow{E}, + \rangle$ consisting of a nonempty set E (of *points*), a vector space $\overrightarrow{E}$ (of *translations*, or *free vectors*), and an action $+: E \times \overrightarrow{E} \to E$, satisfying the following conditions.

(A1) $a + 0 = a$, for every $a \in E$.

(A2) $(a + u) + v = a + (u + v)$, for every $a \in E$, and every $u, v \in \overrightarrow{E}$.

(A3) For any two points $a, b \in E$, there is a unique $u \in \overrightarrow{E}$ such that $a + u = b$.

The unique vector $u \in \overrightarrow{E}$ such that $a + u = b$ is denoted by $\overrightarrow{ab}$, or sometimes by $\mathbf{ab}$, or even by $b - a$. Thus, we also write

$$b = a + \overrightarrow{ab}$$

(or $b = a + \mathbf{ab}$, or even $b = a + (b - a)$).

It is important to note that *adding or rescaling points does not make sense*! However, using the fact that $\overrightarrow{E}$ acts on E is a special way (this action is transitive and faithful), it is possible to define rigorously the notion of *affine combinations* of points and to define affine spaces, affine maps, *etc.* However, this would lead us to far afield, and for our purposes it is enough to stick to vector spaces and we will not distinguish between vector addition $+$ and translation of a point by a vector $+$. Still, one should be aware that affine combinations really apply to points, and that points are not vectors!

If E and F are finite dimensional vector spaces with $\dim(E) = n$ and $\dim(F) = m$, then it is useful to represent an affine map with respect to bases in E in F. However, the translation part c of the affine map must be somehow incorporated. There is a standard trick to do this which amounts to viewing an affine map as a linear map between spaces of dimension $n+1$ and $m+1$. We also have the extra flexibility of choosing origins $a \in E$ and $b \in F$.

Let $(u_1, \ldots, u_n)$ be a basis of E, $(v_1, \ldots, v_m)$ be a basis of F, and let $a \in E$ and $b \in F$ be any two fixed vectors viewed as *origins*. Our affine map f has the property that if $v = f(u)$, then

$$v - b = f(a + u - a) - b = f(a) - b + h(u - a),$$

where the last equality made use of the fact that $h(x) = f(a+x) - f(a)$. If we let $y = v - b$, $x = u - a$, and $d = f(a) - b$, then

$$y = h(x) + d, \quad x \in E.$$

Over the basis $\mathcal{U} = (u_1, \ldots, u_n)$, we write

$$x = x_1 u_1 + \cdots + x_n u_n,$$

and over the basis $\mathcal{V} = (v_1, \ldots, v_m)$, we write

$$\begin{aligned} y &= y_1 v_1 + \cdots + y_m v_m, \\ d &= d_1 v_1 + \cdots + d_m v_m. \end{aligned}$$

Then since

$$y = h(x) + d,$$

if we let A be the $m \times n$ matrix representing the linear map h, that is, the jth column of A consists of the coordinates of $h(u_j)$ over the basis $(v_1, \ldots, v_m)$, then we can write

$$y_{\mathcal{V}} = A x_{\mathcal{U}} + d_{\mathcal{V}}.$$

where $x_{\mathcal{U}} = (x_1, \ldots, x_n)^\top$, $y_{\mathcal{V}} = (y_1, \ldots, y_m)^\top$, and $d_{\mathcal{V}} = (d_1, \ldots, d_m)^\top$. The above is the matrix representation of our affine map f with respect to $(a, (u_1, \ldots, u_n))$ and $(b, (v_1, \ldots, v_m))$.

The reason for using the origins a and b is that it gives us more flexibility. In particular, we can choose $b = f(a)$, and then f behaves like a linear map with respect to the origins a and $b = f(a)$.

When $E = F$, if there is some $a \in E$ such that $f(a) = a$ (a is a *fixed point* of f), then we can pick $b = a$. Then because $f(a) = a$, we get

$$v = f(u) = f(a + u - a) = f(a) + h(u - a) = a + h(u - a),$$

that is

$$v - a = h(u - a).$$

With respect to the new origin a, if we define x and y by

$$\begin{aligned} x &= u - a \\ y &= v - a, \end{aligned}$$

then we get

$$y = h(x).$$

Therefore, f really behaves like a linear map, but *with respect to the new origin* a (not the *standard origin* 0). This is the case of a rotation around an axis that does not pass through the origin.

Remark: A pair $(a, (u_1, \ldots, u_n))$ where $(u_1, \ldots, u_n)$ is a basis of E and a is an origin chosen in E is called an *affine frame*.

We now describe the trick which allows us to incorporate the translation part d into the matrix A. We define the $(m+1) \times (n+1)$ matrix A' obtained by first adding d as the $(n+1)$th column and then $(\underbrace{0, \ldots, 0}_{n}, 1)$ as the $(m+1)$th row:

$$A' = \begin{pmatrix} A & d \\ 0_n & 1 \end{pmatrix}.$$

It is clear that

$$\begin{pmatrix} y \\ 1 \end{pmatrix} = \begin{pmatrix} A & d \\ 0_n & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix}$$

iff

$$y = Ax + d.$$

This amounts to considering a point $x \in \mathbb{R}^n$ as a point $(x, 1)$ *in the (affine) hyperplane* H_{n+1} *in* $\mathbb{R}^{n+1}$ *of equation* $x_{n+1} = 1$. Then an affine map is the restriction to the hyperplane H_{n+1} of the linear map $\widehat{f}$ from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{m+1}$ corresponding to the matrix A' which maps H_{n+1} into H_{m+1} ($\widehat{f}(H_{n+1}) \subseteq H_{m+1}$). Figure 5.4 illustrates this process for $n = 2$.

For example, the map

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

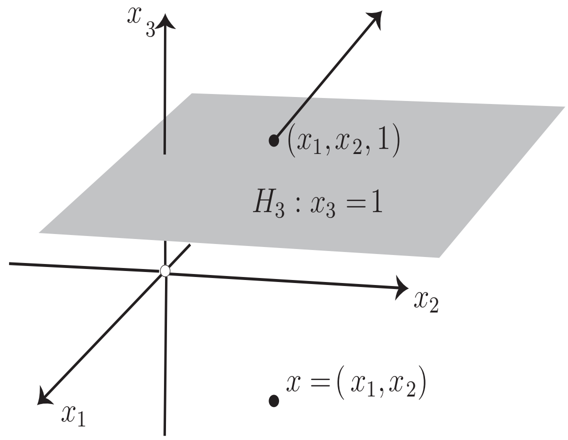

Figure 5.4: Viewing $\mathbb{R}^n$ as a hyperplane in $\mathbb{R}^{n+1}$ ($n = 2$)

defines an affine map f which is represented in $\mathbb{R}^3$ by

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 3 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

It is easy to check that the point $a = (6, -3)$ is fixed by f, which means that $f(a) = a$, so by translating the coordinate frame to the origin a, the affine map behaves like a linear map.

The idea of considering $\mathbb{R}^n$ as an hyperplane in $\mathbb{R}^{n+1}$ can be used to define *projective maps*.

5.5 Summary

The main concepts and results of this chapter are listed below:

- *Direct products, sums, direct sums*.
- *Projections*.
- The fundamental equation

  $$\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(\mathrm{Ker}\, f) + \mathrm{rk}(f)$$

  (The rank-nullity theorem; Theorem 5.8).
- *Grassmann's relation*

  $$\dim(U) + \dim(V) = \dim(U + V) + \dim(U \cap V).$$

- Characterizations of a bijective linear map $f\colon E \to F$.
- *Rank* of a matrix.
- *Affine Maps*.

5.6 Problems

Problem 5.1. Let V and W be two subspaces of a vector space E. Prove that if $V \cup W$ is a subspace of E, then either $V \subseteq W$ or $W \subseteq V$.

Problem 5.2. Prove that for every vector space E, if $f\colon E \to E$ is an idempotent linear map, i.e., $f \circ f = f$, then we have a direct sum

$$E = \operatorname{Ker} f \oplus \operatorname{Im} f,$$

so that f is the projection onto its image $\operatorname{Im} f$.

Problem 5.3. Let $U_1, \ldots, U_p$ be any $p \geq 2$ subspaces of some vector space E and recall that the linear map

$$a\colon U_1 \times \cdots \times U_p \to E$$

is given by

$$a(u_1, \ldots, u_p) = u_1 + \cdots + u_p,$$

with $u_i \in U_i$ for $i = 1, \ldots, p$.

(1) If we let $Z_i \subseteq U_1 \times \cdots \times U_p$ be given by

$$Z_i = \left\{ \left(u_1, \ldots, u_{i-1}, -\sum_{j=1, j\neq i}^{p} u_j, u_{i+1}, \ldots, u_p\right) \;\middle|\; \sum_{j=1, j\neq i}^{p} u_j \in U_i \cap \left(\sum_{j=1, j\neq i}^{p} U_j\right) \right\},$$

for $i = 1, \ldots, p$, then prove that

$$\operatorname{Ker} a = Z_1 = \cdots = Z_p.$$

In general, for any given i, the condition $U_i \cap \left(\sum_{j=1, j\neq i}^{p} U_j\right) = (0)$ does not necessarily imply that $Z_i = (0)$. Thus, let

$$Z = \left\{ \left(u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_p\right) \;\middle|\; u_i = -\sum_{j=1, j\neq i}^{p} u_j,\ u_i \in U_i \cap \left(\sum_{j=1, j\neq i}^{p} U_j\right),\ 1 \leq i \leq p \right\}.$$

Since $\mathrm{Ker}\, a = Z_1 = \cdots = Z_p$, we have $Z = \mathrm{Ker}\, a$. Prove that if

$$U_i \cap \left(\sum_{j=1, j\neq i}^{p} U_j \right) = (0) \quad 1 \leq i \leq p,$$

then $Z = \mathrm{Ker}\, a = (0)$.

(2) Prove that $U_1 + \cdots + U_p$ is a direct sum iff

$$U_i \cap \left(\sum_{j=1, j\neq i}^{p} U_j \right) = (0) \quad 1 \leq i \leq p.$$

Problem 5.4. Assume that E is finite-dimensional, and let $f_i \colon E \to E$ be any $p \geq 2$ linear maps such that

$$f_1 + \cdots + f_p = \mathrm{id}_E.$$

Prove that the following properties are equivalent:

(1) $f_i^2 = f_i$, $1 \leq i \leq p$.

(2) $f_j \circ f_i = 0$, for all $i \neq j$, $1 \leq i, j \leq p$.

Hint. Use Problem 5.2.

Let $U_1, \ldots, U_p$ be any $p \geq 2$ subspaces of some vector space E. Prove that $U_1 + \cdots + U_p$ is a direct sum iff

$$U_i \cap \left(\sum_{j=1}^{i-1} U_j \right) = (0), \quad i = 2, \ldots, p.$$

Problem 5.5. Given any vector space E, a linear map $f \colon E \to E$ is an *involution* if $f \circ f = \mathrm{id}$.

(1) Prove that an involution f is invertible. What is its inverse?

(2) Let E_1 and E_{-1} be the subspaces of E defined as follows:

$$\begin{aligned} E_1 &= \{u \in E \mid f(u) = u\} \\ E_{-1} &= \{u \in E \mid f(u) = -u\}. \end{aligned}$$

Prove that we have a direct sum

$$E = E_1 \oplus E_{-1}.$$

Hint. For every $u \in E$, write

$$u = \frac{u + f(u)}{2} + \frac{u - f(u)}{2}.$$

(3) If E is finite-dimensional and f is an involution, prove that there is some basis of E with respect to which the matrix of f is of the form

$$I_{k,n-k} = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix},$$

where I_k is the $k \times k$ identity matrix (similarly for I_{n-k}) and $k = \dim(E_1)$. Can you give a geometric interpretation of the action of f (especially when $k = n-1$)?

Problem 5.6. An $n \times n$ matrix H is *upper Hessenberg* if $h_{jk} = 0$ for all (j,k) such that $j - k \geq 0$. An upper Hessenberg matrix is *unreduced* if $h_{i+1i} \neq 0$ for $i = 1, \ldots, n-1$.

Prove that if H is a singular unreduced upper Hessenberg matrix, then $\dim(\mathrm{Ker}\,(H)) = 1$.

Problem 5.7. Let A be any $n \times k$ matrix.

(1) Prove that the $k \times k$ matrix $A^\top A$ and the matrix A have the same nullspace. Use this to prove that $\mathrm{rank}(A^\top A) = \mathrm{rank}(A)$. Similarly, prove that the $n \times n$ matrix $AA^\top$ and the matrix $A^\top$ have the same nullspace, and conclude that $\mathrm{rank}(AA^\top) = \mathrm{rank}(A^\top)$.

We will prove later that $\mathrm{rank}(A^\top) = \mathrm{rank}(A)$.

(2) Let $a_1, \ldots, a_k$ be k linearly independent vectors in $\mathbb{R}^n$ $(1 \leq k \leq n)$, and let A be the $n \times k$ matrix whose ith column is a_i. Prove that $A^\top A$ has rank k, and that it is invertible. Let $P = A(A^\top A)^{-1}A^\top$ (an $n \times n$ matrix). Prove that

$$\begin{aligned} P^2 &= P \\ P^\top &= P. \end{aligned}$$

What is the matrix P when $k = 1$?

(3) Prove that the image of P is the subspace V spanned by $a_1, \ldots, a_k$, or equivalently the set of all vectors in $\mathbb{R}^n$ of the form Ax, with $x \in \mathbb{R}^k$. Prove that the nullspace U of P is the set of vectors $u \in \mathbb{R}^n$ such that $A^\top u = 0$. Can you give a geometric interpretation of U?

Conclude that P is a projection of $\mathbb{R}^n$ onto the subspace V spanned by $a_1, \ldots, a_k$, and that

$$\mathbb{R}^n = U \oplus V.$$

Problem 5.8. A *rotation* R_θ in the plane $\mathbb{R}^2$ is given by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

(1) Use `Matlab` to show the action of a rotation R_θ on a simple figure such as a triangle or a rectangle, for various values of θ, including $\theta = \pi/6, \pi/4, \pi/3, \pi/2$.

(2) Prove that R_θ is invertible and that its inverse is $R_{-\theta}$.

(3) For any two rotations R_α and R_β, prove that

$$R_\beta \circ R_\alpha = R_\alpha \circ R_\beta = R_{\alpha+\beta}.$$

Use (2)-(3) to prove that the rotations in the plane form a commutative group denoted $\mathbf{SO}(2)$.

Problem 5.9. Consider the affine map $R_{\theta,(a_1,a_2)}$ in $\mathbb{R}^2$ given by

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

(1) Prove that if $\theta \neq k2\pi$, with $k \in \mathbb{Z}$, then $R_{\theta,(a_1,a_2)}$ has a unique fixed point (c_1, c_2), that is, there is a unique point (c_1, c_2) such that

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = R_{\theta,(a_1,a_2)} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

and this fixed point is given by

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{2\sin(\theta/2)} \begin{pmatrix} \cos(\pi/2 - \theta/2) & -\sin(\pi/2 - \theta/2) \\ \sin(\pi/2 - \theta/2) & \cos(\pi/2 - \theta/2) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

(2) In this question we still assume that $\theta \neq k2\pi$, with $k \in \mathbb{Z}$. By translating the coordinate system with origin $(0,0)$ to the new coordinate system with origin (c_1, c_2), which means that if (x_1, x_2) are the coordinates with respect to the standard origin $(0,0)$ and if (x_1', x_2') are the coordinates with respect to the new origin (c_1, c_2), we have

$$\begin{aligned} x_1 &= x_1' + c_1 \\ x_2 &= x_2' + c_2 \end{aligned}$$

and similarly for (y_1, y_2) and (y_1', y_2'), then show that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = R_{\theta,(a_1,a_2)} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

becomes

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = R_\theta \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}.$$

Conclude that with respect to the new origin (c_1, c_2), the affine map $R_{\theta,(a_1,a_2)}$ becomes the rotation R_θ. We say that $R_{\theta,(a_1,a_2)}$ is a *rotation of center* (c_1, c_2).

(3) Use `Matlab` to show the action of the affine map $R_{\theta,(a_1,a_2)}$ on a simple figure such as a triangle or a rectangle, for $\theta = \pi/3$ and various values of (a_1, a_2). Display the center (c_1, c_2) of the rotation.

What kind of transformations correspond to $\theta = k2\pi$, with $k \in \mathbb{Z}$?

(4) Prove that the inverse of $R_{\theta,(a_1,a_2)}$ is of the form $R_{-\theta,(b_1,b_2)}$, and find (b_1, b_2) in terms of θ and (a_1, a_2).

(5) Given two affine maps $R_{\alpha,(a_1,a_2)}$ and $R_{\beta,(b_1,b_2)}$, prove that

$$R_{\beta,(b_1,b_2)} \circ R_{\alpha,(a_1,a_2)} = R_{\alpha+\beta,(t_1,t_2)}$$

for some (t_1, t_2), and find (t_1, t_2) in terms of β, (a_1, a_2) and (b_1, b_2).

Even in the case where $(a_1, a_2) = (0, 0)$, prove that in general

$$R_{\beta,(b_1,b_2)} \circ R_\alpha \neq R_\alpha \circ R_{\beta,(b_1,b_2)}.$$

Use (4)-(5) to show that the affine maps of the plane defined in this problem form a nonabelian group denoted $\mathbf{SE}(2)$.

Prove that $R_{\beta,(b_1,b_2)} \circ R_{\alpha,(a_1,a_2)}$ is not a translation (possibly the identity) iff $\alpha + \beta \neq k2\pi$, for all $k \in \mathbb{Z}$. Find its center of rotation when $(a_1, a_2) = (0, 0)$.

If $\alpha + \beta = k2\pi$, then $R_{\beta,(b_1,b_2)} \circ R_{\alpha,(a_1,a_2)}$ is a pure translation. Find the translation vector of $R_{\beta,(b_1,b_2)} \circ R_{\alpha,(a_1,a_2)}$.

Problem 5.10. (Affine subspaces) A subset $\mathcal{A}$ of $\mathbb{R}^n$ is called an *affine subspace* if either $\mathcal{A} = \emptyset$, or there is some vector $a \in \mathbb{R}^n$ and some subspace U of $\mathbb{R}^n$ such that

$$\mathcal{A} = a + U = \{a + u \mid u \in U\}.$$

We define the dimension $\dim(\mathcal{A})$ of $\mathcal{A}$ as the dimension $\dim(U)$ of U.

(1) If $\mathcal{A} = a + U$, why is $a \in \mathcal{A}$?

What are affine subspaces of dimension 0? What are affine subspaces of dimension 1 (begin with $\mathbb{R}^2$)? What are affine subspaces of dimension 2 (begin with $\mathbb{R}^3$)?

Prove that any nonempty affine subspace is closed under affine combinations.

(2) Prove that if $\mathcal{A} = a + U$ is any nonempty affine subspace, then $\mathcal{A} = b + U$ for any $b \in \mathcal{A}$.

(3) Let $\mathcal{A}$ be any nonempty subset of $\mathbb{R}^n$ closed under affine combinations. For any $a \in \mathcal{A}$, prove that

$$U_a = \{x - a \in \mathbb{R}^n \mid x \in \mathcal{A}\}$$

is a (linear) subspace of $\mathbb{R}^n$ such that

$$\mathcal{A} = a + U_a.$$

Prove that U_a does not depend on the choice of $a \in \mathcal{A}$; that is, $U_a = U_b$ for all $a, b \in \mathcal{A}$. In fact, prove that

$$U_a = U = \{y - x \in \mathbb{R}^n \mid x, y \in \mathcal{A}\}, \quad \text{for all } a \in \mathcal{A},$$

and so

$$\mathcal{A} = a + U, \quad \text{for any } a \in \mathcal{A}.$$

Remark: The subspace U is called the *direction* of $\mathcal{A}$.

(4) Two nonempty affine subspaces $\mathcal{A}$ and $\mathcal{B}$ are said to be *parallel* iff they have the same direction. Prove that that if $\mathcal{A} \neq \mathcal{B}$ and $\mathcal{A}$ and $\mathcal{B}$ are parallel, then $\mathcal{A} \cap \mathcal{B} = \emptyset$.

Remark: The above shows that affine subspaces behave quite differently from linear subspaces.

Problem 5.11. (Affine frames and affine maps) For any vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, let $\widehat{v} \in \mathbb{R}^{n+1}$ be the vector $\widehat{v} = (v_1, \ldots, v_n, 1)$. Equivalently, $\widehat{v} = (\widehat{v}_1, \ldots, \widehat{v}_{n+1}) \in \mathbb{R}^{n+1}$ is the vector defined by

$$\widehat{v}_i = \begin{cases} v_i & \text{if } 1 \leq i \leq n, \\ 1 & \text{if } i = n+1. \end{cases}$$

(1) For any $m+1$ vectors $(u_0, u_1, \ldots, u_m)$ with $u_i \in \mathbb{R}^n$ and $m \leq n$, prove that if the m vectors $(u_1 - u_0, \ldots, u_m - u_0)$ are linearly independent, then the $m+1$ vectors $(\widehat{u}_0, \ldots, \widehat{u}_m)$ are linearly independent.

(2) Prove that if the $m+1$ vectors $(\widehat{u}_0, \ldots, \widehat{u}_m)$ are linearly independent, then for any choice of i, with $0 \leq i \leq m$, the m vectors $u_j - u_i$ for $j \in \{0, \ldots, m\}$ with $j - i \neq 0$ are linearly independent.

Any $m+1$ vectors $(u_0, u_1, \ldots, u_m)$ such that the $m+1$ vectors $(\widehat{u}_0, \ldots, \widehat{u}_m)$ are linearly independent are said to be *affinely independent*.

From (1) and (2), the vector $(u_0, u_1, \ldots, u_m)$ are affinely independent iff for any any choice of i, with $0 \leq i \leq m$, the m vectors $u_j - u_i$ for $j \in \{0, \ldots, m\}$ with $j - i \neq 0$ are linearly independent. If $m = n$, we say that $n+1$ affinely independent vectors $(u_0, u_1, \ldots, u_n)$ form an *affine frame* of $\mathbb{R}^n$.

(3) if $(u_0, u_1, \ldots, u_n)$ is an affine frame of $\mathbb{R}^n$, then prove that for every vector $v \in \mathbb{R}^n$, there is a unique $(n+1)$-tuple $(\lambda_0, \lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n+1}$, with $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$, such that

$$v = \lambda_0 u_0 + \lambda_1 u_1 + \cdots + \lambda_n u_n.$$

The scalars $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ are called the *barycentric* (or *affine*) *coordinates* of v w.r.t. the affine frame $(u_0, u_1, \ldots, u_n)$.

If we write $e_i = u_i - u_0$, for $i = 1, \ldots, n$, then prove that we have

$$v = u_0 + \lambda_1 e_1 + \cdots + \lambda_n e_n,$$

and since $(e_1, \ldots, e_n)$ is a basis of $\mathbb{R}^n$ (by (1) & (2)), the n-tuple $(\lambda_1, \ldots, \lambda_n)$ consists of the standard coordinates of $v - u_0$ over the basis $(e_1, \ldots, e_n)$.

Conversely, for any vector $u_0 \in \mathbb{R}^n$ and for any basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$, let $u_i = u_0 + e_i$ for $i = 1, \ldots, n$. Prove that $(u_0, u_1, \ldots, u_n)$ is an affine frame of $\mathbb{R}^n$, and for any $v \in \mathbb{R}^n$, if

$$v = u_0 + x_1 e_1 + \cdots + x_n e_n,$$

with $(x_1, \ldots, x_n) \in \mathbb{R}^n$ (unique), then

$$v = (1 - (x_1 + \cdots + x_x))u_0 + x_1 u_1 + \cdots + x_n u_n,$$

so that $(1 - (x_1 + \cdots + x_x)), x_1, \cdots, x_n)$, are the barycentric coordinates of v w.r.t. the affine frame $(u_0, u_1, \ldots, u_n)$.

The above shows that there is a one-to-one correspondence between affine frames $(u_0, \ldots, u_n)$ and pairs $(u_0, (e_1, \ldots, e_n))$, with $(e_1, \ldots, e_n)$ a basis. Given an affine frame $(u_0, \ldots, u_n)$, we obtain the basis $(e_1, \ldots, e_n)$ with $e_i = u_i - u_0$, for $i = 1, \ldots, n$; given the pair $(u_0, (e_1, \ldots, e_n))$ where $(e_1, \ldots, e_n)$ is a basis, we obtain the affine frame $(u_0, \ldots, u_n)$, with $u_i = u_0 + e_i$, for $i = 1, \ldots, n$. There is also a one-to-one correspondence between barycentric coordinates w.r.t. the affine frame $(u_0, \ldots, u_n)$ and standard coordinates w.r.t. the basis $(e_1, \ldots, e_n)$. The barycentric cordinates $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ of v (with $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$) yield the standard coordinates $(\lambda_1, \ldots, \lambda_n)$ of $v - u_0$; the standard coordinates $(x_1, \ldots, x_n)$ of $v - u_0$ yield the barycentric coordinates $(1 - (x_1 + \cdots + x_n), x_1, \ldots, x_n)$ of v.

(4) Let $(u_0, \ldots, u_n)$ be any affine frame in $\mathbb{R}^n$ and let $(v_0, \ldots, v_n)$ be any vectors in $\mathbb{R}^m$. Prove that there is a *unique* affine map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(u_i) = v_i, \quad i = 0, \ldots, n.$$

(5) Let $(a_0, \ldots, a_n)$ be any affine frame in $\mathbb{R}^n$ and let $(b_0, \ldots, b_n)$ be any $n+1$ points in $\mathbb{R}^n$. Prove that there is a unique $(n+1) \times (n+1)$ matrix

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

corresponding to the unique affine map f such that

$$f(a_i) = b_i, \quad i = 0, \ldots, n,$$

in the sense that

$$A\widehat{a}_i = \widehat{b}_i, \quad i = 0, \ldots, n,$$

and that A is given by

$$A = \begin{pmatrix} \widehat{b}_0 & \widehat{b}_1 & \cdots & \widehat{b}_n \end{pmatrix} \begin{pmatrix} \widehat{a}_0 & \widehat{a}_1 & \cdots & \widehat{a}_n \end{pmatrix}^{-1}.$$

Make sure to prove that the bottom row of A is $(0, \ldots, 0, 1)$.

In the special case where $(a_0, \ldots, a_n)$ is the canonical affine frame with $a_i = e_{i+1}$ for $i = 0, \ldots, n-1$ and $a_n = (0, \ldots, 0)$ (where e_i is the ith canonical basis vector), show that

$$\begin{pmatrix} \widehat{a}_0 & \widehat{a}_1 & \cdots & \widehat{a}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} \widehat{a}_0 & \widehat{a}_1 & \cdots & \widehat{a}_n \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & \cdots & -1 & 1 \end{pmatrix}.$$

For example, when $n = 2$, if we write $b_i = (x_i, y_i)$, then we have

$$A = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} x_1 - x_3 & x_2 - x_3 & x_3 \\ y_1 - y_3 & y_2 - y_3 & y_3 \\ 0 & 0 & 1 \end{pmatrix}.$$

(6) Recall that a nonempty affine subspace $\mathcal{A}$ of $\mathbb{R}^n$ is any nonempty subset of $\mathbb{R}^n$ closed under affine combinations. For any affine map $f\colon \mathbb{R}^n \to \mathbb{R}^m$, for any affine subspace $\mathcal{A}$ of $\mathbb{R}^n$, and any affine subspace $\mathcal{B}$ of $\mathbb{R}^m$, prove that $f(\mathcal{A})$ is an affine subspace of $\mathbb{R}^m$, and that $f^{-1}(\mathcal{B})$ is an affine subspace of $\mathbb{R}^n$.