Chapter 3

Direct Sums, Affine Maps, The Dual Space, Duality

3.1 Direct Products, Sums, and Direct Sums

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

Definition 3.1. 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 $\lambda \in \mathbb{R}$.

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 $E_1, \ldots, E_p$.

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.

It can be shown (using bases) that

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

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 $U_1, \ldots, U_p$.

If the map a is injective, then $\mathrm{Ker}\, a = \{(\underbrace{0, \ldots, 0}_{p})\}$, which means that if $u_i \in U_i$ for $i = 1, \ldots, p$ and if

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

then $u_1 = 0, \ldots, u_p = 0$.

In this case, 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 $i = 1, \ldots, p$.

It is also clear that for any p nonzero vectors $u_i \in U_i$, $u_1, \ldots, u_p$ are linearly independent.

Definition 3.2. For any vector space E and any $p \geq 2$ subspaces $U_1, \ldots, U_p$ of E, if the map a 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

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

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.

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

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

which implies that

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

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 $\operatorname{Ker} a$, so $\operatorname{Ker} a$ is isomorphic to $U_1 \cap U_2$.

As a consequence, we get the following result:

Proposition 3.1. *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 $U_1 \cap U_2 = (0)$.*

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

$$\begin{aligned}\mathbf{S}(n) &= \{A \in \mathrm{M}_n \mid A^\top = A\}\\ \mathbf{Skew}(n) &= \{A \in \mathrm{M}_n \mid A^\top = -A\}\end{aligned}$$

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 have the direct sum

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

Proposition 3.2. *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

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

The isomorphism $U_1 \times \cdots \times U_p \approx U_1 \oplus \cdots \oplus U_p$ implies

Proposition 3.3. *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

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

for some $u_i \in U_i$ 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 3.4. *Let E be a vector space. For any $p \geq 2$ linear maps $f_i \colon E \to E$, 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 3.5. *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*

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

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.

Theorem 3.6. *Let $f\colon E \to F$ be a linear map. 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 $s\colon \mathrm{Im}\, f \to E$ 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:

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

As a consequence,

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

Remark: The dimension $\dim(\mathrm{Ker}\, f)$ of the kernel of a linear map f is often called the *nullity* of f.

We now derive some important results using Theorem 3.6.

Proposition 3.7. *Given a vector space E, if U and V are any two subspaces of E, then*

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

an equation known as Grassmann's relation.

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 $U \cap V = (0)$.

However, we can show that if $\dim(U) = 3$ and $\dim(V) = 3$, then $\dim(U \cap V) \geq 1$.

As another consequence of Proposition 3.7, if U and V are two hyperplanes in a vector space of dimension n, so that $\dim(U) = n - 1$ and $\dim(V) = n - 1$, we have

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

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

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

Proposition 3.8. *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 $U_1 \oplus \cdots \oplus U_p$.

Another important corollary of Theorem 3.6 is the following result:

Proposition 3.9. *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) $\mathrm{Ker}\, f = (0)$.

One should be warned that Proposition 3.9 fails in infinite dimension.

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

Proposition 3.10. *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 $r \circ f = \mathrm{id}_E$. 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 $f \circ s = \mathrm{id}_F$.*

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.

Proposition 3.11. *Given a linear map $f\colon E \to F$, the following properties hold:*

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

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

The rank of a matrix is defined as follows.

Definition 3.3. 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 1.4, 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.

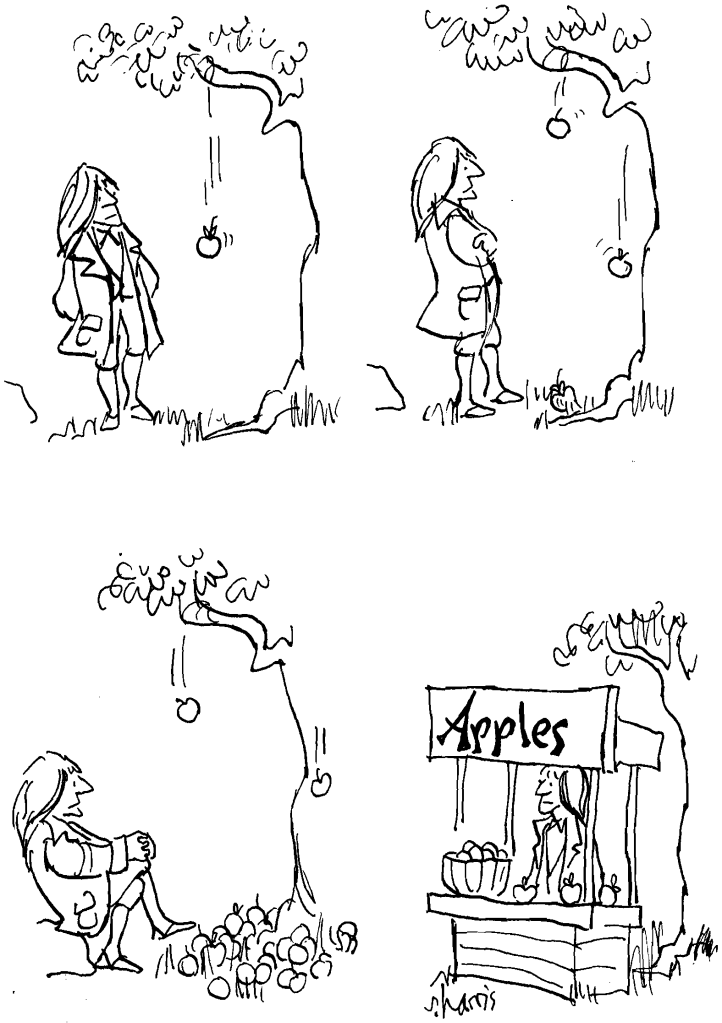

Figure 3.1: How did Newton start a business

3.2 Affine Maps

We showed in Section 1.5 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 pratice (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 1.5 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 0 with the coefficient $1 - \sum_{i\in I} \lambda_i$.

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 3.12. *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.

Surprisingly, the converse of Proposition 3.12 also holds.

Proposition 3.13. *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*

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

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

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

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

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

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 3.13 shows that the definition of h does not depend on the origin chosen in E. Also, since

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

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 $\overrightarrow{f}$.

Observe that the linear map associated with a pure translation is the identity.

In view of Propositions 3.12 and 3.13, it is natural to make the following definition.

Definition 3.4. For any two vector spaces E and F, a function $f\colon E \to F$ 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 $f\colon E \to F$ is an *affine map* if there is some linear map $h\colon E \to F$ (also denoted by $\overrightarrow{f}$) and some fixed vector $c \in F$ such that

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

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*, but this would lead us to far afield, and for our purposes, it is enough to stick to vector spaces.

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 somewhow 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

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

So, 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 $d_{\mathcal{V}} = (d_1, \ldots, d_m)^\top$.

This is the *matrix representation* of our affine map f.

The reason for using the origins a and b is that it gives us more flexibility.

In particular, 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

$$y = h(x).$$

Then, 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:

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

Then, 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 3.2 illustrates this process for $n = 2$.

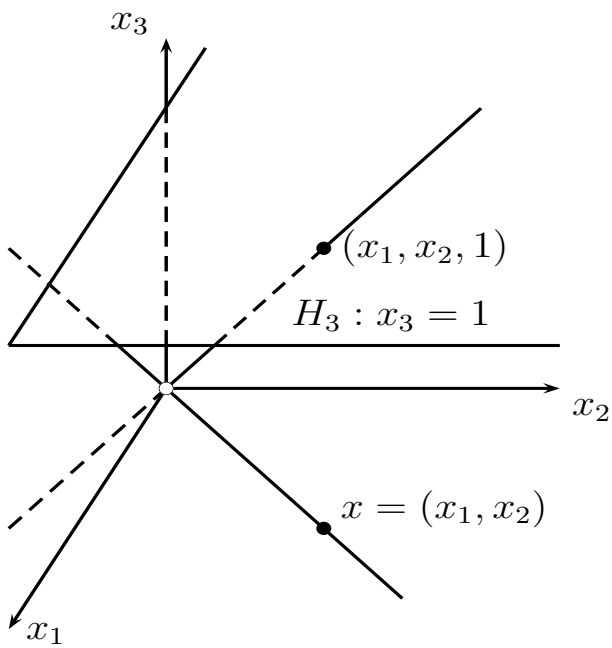

Figure 3.2: Viewing $\mathbb{R}^n$ as a hyperplane in $\mathbb{R}^{n+1}$ ($n = 2$)

For example, the map

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

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*.

Figure 3.3: Dog Logic

3.3 The Dual Space E^* and Linear Forms

We already observed that the field K itself ($K = \mathbb{R}$ or $K = \mathbb{C}$) is a vector space (over itself).

The vector space $\mathrm{Hom}(E, K)$ of linear maps from E to the field K, the *linear forms*, plays a particular role.

We take a quick look at the connection between E and $E^* = \mathrm{Hom}(E, K)$, its *dual space*.

As we will see shortly, every linear map $f\colon E \to F$ gives rise to a linear map $f^\top\colon F^* \to E^*$, and it turns out that in a suitable basis, the matrix of $f^\top$ is the *transpose* of the matrix of f.

Thus, the notion of dual space provides a conceptual explanation of the phenomena associated with transposition.

But it does more, because it allows us to view subspaces as solutions of sets of linear equations and vice-versa.

Consider the following set of two "linear equations" in $\mathbb{R}^3$,

$$\begin{aligned} x - y + z &= 0 \\ x - y - z &= 0, \end{aligned}$$

and let us find out what is their set V of common solutions $(x, y, z) \in \mathbb{R}^3$.

By subtracting the second equation from the first, we get $2z = 0$, and by adding the two equations, we find that $2(x - y) = 0$, so the set V of solutions is given by

$$\begin{aligned} y &= x \\ z &= 0. \end{aligned}$$

This is a one dimensional subspace of $\mathbb{R}^3$. Geometrically, this is the line of equation $y = x$ in the plane $z = 0$.

Now, why did we say that the above equations are linear?

This is because, as functions of (x, y, z), both maps $f_1 \colon (x, y, z) \mapsto x - y + z$ and $f_2 \colon (x, y, z) \mapsto x - y - z$ are linear.

The set of all such linear functions from $\mathbb{R}^3$ to $\mathbb{R}$ is a vector space; we used this fact to form linear combinations of the "equations" f_1 and f_2.

Observe that the dimension of the subspace V is 1.

The ambient space has dimension $n = 3$ and there are two "independent" equations f_1, f_2, so it appears that the dimension $\dim(V)$ of the subspace V defined by m independent equations is

$$\dim(V) = n - m,$$

which is indeed a general fact (proved in Theorem 3.14).

More generally, in $\mathbb{R}^n$, a linear equation is determined by an n-tuple $(a_1, \ldots, a_n) \in \mathbb{R}^n$, and the solutions of this linear equation are given by the n-tuples $(x_1, \ldots, x_n) \in \mathbb{R}^n$ such that

$$a_1 x_1 + \cdots + a_n x_n = 0;$$

these solutions constitute the kernel of the linear map $(x_1, \ldots, x_n) \mapsto a_1 x_1 + \cdots + a_n x_n$.

The above considerations assume that we are working in the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$, but we can define "linear equations" independently of bases and in any dimension, by viewing them as elements of the vector space $\mathrm{Hom}(E, K)$ of linear maps from E to the field K.

Definition 3.5. Given a vector space E, the vector space $\mathrm{Hom}(E, K)$ of linear maps from E to K is called the *dual space (or dual)* of E. The space $\mathrm{Hom}(E, K)$ is also denoted by E^*, and the linear maps in E^* are called *the linear forms*, or *covectors*. The dual space E^{**} of the space E^* is called the *bidual* of E.

As a matter of notation, linear forms $f\colon E \to K$ will also be denoted by starred symbol, such as u^*, x^*, etc.

If E is a vector space of finite dimension n and $(u_1, \ldots, u_n)$ is a basis of E, for any linear form $f^* \in E^*$, for every $x = x_1u_1 + \cdots + x_nu_n \in E$, by linearity we have

$$\begin{aligned} f^*(x) &= f^*(u_1)x_1 + \cdots + f^*(n_n)x_n \\ &= \lambda_1 x_1 + \cdots + \lambda_n x_n, \end{aligned}$$

with $\lambda_i = f^*(u_i) \in K$ for every i, $1 \leq i \leq n$.

Thus, with respect to the basis $(u_1, \ldots, u_n)$, the linear form f^* is represented by the row vector

$$(\lambda_1 \ \cdots \ \lambda_n),$$

we have

$$f^*(x) = \begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

a linear combination of the coordinates of x, and we can view the linear form f^* as a *linear equation*.

If we decide to use a column vector of coefficients

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

instead of a row vector, then the linear form f^* is defined by

$$f^*(x) = c^\top x.$$

The above notation is often used in machine learning.

Example 3.1. Given any differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$, by definition, for any $x \in \mathbb{R}^n$, the *total derivative* df_x of f at x is the linear form $df_x\colon \mathbb{R}^n \to \mathbb{R}$ defined so that for all $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$,

$$df_x(u) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) & \cdots & \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, u_i.$$

Example 3.2. Let $\mathcal{C}([0,1])$ be the vector space of continuous functions $f\colon [0,1] \to \mathbb{R}$. The map $\mathcal{I}\colon \mathcal{C}([0,1]) \to \mathbb{R}$ given by

$$\mathcal{I}(f) = \int_0^1 f(x)dx \quad \text{for any } f \in \mathcal{C}([0,1])$$

is a linear form (integration).

Example 3.3. Consider the vector space $\mathrm{M}_n(\mathbb{R})$ of real $n \times n$ matrices. Let $\mathrm{tr}\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ be the function given by

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn},$$

called the *trace* of A. It is a linear form.

Let $s\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$ be the function given by

$$s(A) = \sum_{i,j=1}^{n} a_{ij},$$

where $A = (a_{ij})$. It is immediately verified that s is a linear form.

Given a linear form $u^* \in E^*$ and a vector $v \in E$, the result $u^*(v)$ of applying u^* to v is also denoted by $\langle u^*, v\rangle$.

This defines a binary operation $\langle -, -\rangle \colon E^* \times E \to K$ satisfying the following properties:

$$\begin{aligned}
\langle u_1^* + u_2^*, v\rangle &= \langle u_1^*, v\rangle + \langle u_2^*, v\rangle \\
\langle u^*, v_1 + v_2\rangle &= \langle u^*, v_1\rangle + \langle u^*, v_2\rangle \\
\langle \lambda u^*, v\rangle &= \lambda\langle u^*, v\rangle \\
\langle u^*, \lambda v\rangle &= \lambda\langle u^*, v\rangle.
\end{aligned}$$

The above identities mean that $\langle -, -\rangle$ is a *bilinear map*, since it is linear in each argument.

It is often called the *canonical pairing* between E^* and E.

In view of the above identities, given any fixed vector $v \in E$, the map $\mathrm{eval}_v \colon E^* \to K$ (*evaluation at v*) defined such that

$$\mathrm{eval}_v(u^*) = \langle u^*, v \rangle = u^*(v) \quad \text{for every } u^* \in E^*$$

is a linear map from E^* to K, that is, eval_v is a linear form in E^{**}.

Again from the above identities, the map
$\mathrm{eval}_E \colon E \to E^{**}$, defined such that

$$\mathrm{eval}_E(v) = \mathrm{eval}_v \quad \text{for every } v \in E,$$

is a linear map.

We shall see that it is injective, and that it is an isomorphism when E has finite dimension.

We now formalize the notion of the set V^0 of linear equations vanishing on all vectors in a given subspace $V \subseteq E$, and the notion of the set U^0 of common solutions of a given set $U \subseteq E^*$ of linear equations.

The duality theorem (Theorem 3.14) shows that the dimensions of V and V^0, and the dimensions of U and U^0, are related in a crucial way.

It also shows that, in finite dimension, the maps $V \mapsto V^0$ and $U \mapsto U^0$ are inverse bijections from subspaces of E to subspaces of E^*.

Definition 3.6. Given a vector space E and its dual E^*, we say that a vector $v \in E$ and a linear form $u^* \in E^*$ are *orthogonal* iff $\langle u^*, v\rangle = 0$. Given a subspace V of E and a subspace U of E^*, we say that *V and U are orthogonal* iff $\langle u^*, v\rangle = 0$ for every $u^* \in U$ and every $v \in V$. Given a subset V of E (resp. a subset U of E^*), the *orthogonal V^0 of V* is the subspace V^0 of E^* defined such that

$$V^0 = \{u^* \in E^* \mid \langle u^*, v\rangle = 0, \text{ for every } v \in V\}$$

(resp. the *orthogonal U^0 of U* is the subspace U^0 of E defined such that

$$U^0 = \{v \in E \mid \langle u^*, v\rangle = 0, \text{ for every } u^* \in U\}).$$

The subspace $V^0 \subseteq E^*$ is also called the *annihilator* of V.

The subspace $U^0 \subseteq E$ annihilated by $U \subseteq E^*$ does not have a special name. It seems reasonable to call it the *linear subspace (or linear variety) defined by U*.

Informally, V^0 is the *set of linear equations that vanish on V*, and U^0 is *the set of common zeros of all linear equations in U*. We can also define V^0 by

$$V^0 = \{u^* \in E^* \mid V \subseteq \operatorname{Ker} u^*\}$$

and U^0 by

$$U^0 = \bigcap_{u^* \in U} \operatorname{Ker} u^*.$$

Observe that $E^0 = \{0\} = (0)$, and $\{0\}^0 = E^*$.

Furthermore, if $V_1 \subseteq V_2 \subseteq E$, then $V_2^0 \subseteq V_1^0 \subseteq E^*$, and if $U_1 \subseteq U_2 \subseteq E^*$, then $U_2^0 \subseteq U_1^0 \subseteq E$.

It can also be shown that that $V \subseteq V^{00}$ for every subspace V of E, and that $U \subseteq U^{00}$ for every subspace U of E^*.

We will see shortly that in finite dimension, we have

$$V = V^{00} \quad \text{and} \quad U = U^{00}.$$

Here are some examples. Let $E = \mathrm{M}_2(\mathbb{R})$, the space of real 2×2 matrices, and let V be the subspace of $\mathrm{M}_2(\mathbb{R})$ spanned by the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We check immediately that the subspace V consists of all matrices of the form

$$\begin{pmatrix} b & a \\ a & c \end{pmatrix},$$

that is, all symmetric matrices.

The matrices

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

in V satisfy the equation

$$a_{12} - a_{21} = 0,$$

and all scalar multiples of these equations, so V^0 is the subspace of E^* spanned by the linear form given by

$$u^*(a_{11}, a_{12}, a_{21}, a_{22}) = a_{12} - a_{21}.$$

By the duality theorem (Theorem 3.14) we have

$$\dim(V^0) = \dim(E) - \dim(V) = 4 - 3 = 1.$$

The above example generalizes to $E = \mathrm{M}_n(\mathbb{R})$ for any $n \geq 1$, but this time, consider the space U of linear forms asserting that a matrix A is symmetric; these are the linear forms spanned by the $n(n-1)/2$ equations

$$a_{ij} - a_{ji} = 0, \quad 1 \leq i < j \leq n;$$

Note there are no constraints on diagonal entries, and half of the equations

$$a_{ij} - a_{ji} = 0, \quad 1 \leq i \neq j \leq n$$

are redundant. It is easy to check that the equations (linear forms) for which $i < j$ are linearly independent.

To be more precise, let U be the space of linear forms in E^* spanned by the linear forms

$$\begin{aligned} u^*_{ij}(a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn}) \\ = a_{ij} - a_{ji}, \quad 1 \leq i < j \leq n. \end{aligned}$$

The dimension of U is $n(n-1)/2$. Then, the set U^0 of common solutions of these equations is the space $\mathbf{S}(n)$ of symmetric matrices.

By the duality theorem (Theorem 3.14), this space has dimension

$$\frac{n(n+1)}{2} = n^2 - \frac{n(n-1)}{2}.$$

If $E = \mathrm{M}_n(\mathbb{R})$, consider the subspace U of linear forms in E^* spanned by the linear forms

$$\begin{aligned} u^*_{ij}(a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn}) \\ = a_{ij} + a_{ji}, \quad 1 \leq i < j \leq n \\ u^*_{ii}(a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn}) \\ = a_{ii}, \quad 1 \leq i \leq n. \end{aligned}$$

It is easy to see that these linear forms are linearly independent, so $\dim(U) = n(n+1)/2$.

The space U^0 of matrices $A \in \mathrm{M}_n(\mathbb{R})$ satifying all of the above equations is clearly the space $\mathbf{Skew}(n)$ of skew-symmetric matrices.

By the duality theorem (Theorem 3.14), the dimension of U^0 is

$$\frac{n(n-1)}{2} = n^2 - \frac{n(n+1)}{2}.$$

For yet another example with $E = \mathrm{M}_n(\mathbb{R})$, for any $A \in \mathrm{M}_n(\mathbb{R})$, consider the linear form in E^* given by

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn},$$

called the *trace* of A.

The subspace U^0 of E consisting of all matrices A such that $\mathrm{tr}(A) = 0$ is a space of dimension $n^2 - 1$.

The dimension equations

$$\begin{aligned} \dim(V) + \dim(V^0) &= \dim(E) \\ \dim(U) + \dim(U^0) &= \dim(E) \end{aligned}$$

are always true (if E is finite-dimensional). This is part of the duality theorem (Theorem 3.14).

In constrast with the previous examples, given a matrix $A \in \mathrm{M}_n(\mathbb{R})$, the equations asserting that $A^\top A = I$ are not linear constraints.

For example, for $n = 2$, we have

$$\begin{aligned} a_{11}^2 + a_{21}^2 &= 1 \\ a_{21}^2 + a_{22}^2 &= 1 \\ a_{11}a_{12} + a_{21}a_{22} &= 0. \end{aligned}$$

Given a vector space E and any basis $(u_i)_{i\in I}$ for E, we can associate to each u_i a linear form $u_i^* \in E^*$, and the u_i^* have some remarkable properties.

Definition 3.7. Given a vector space E and any basis $(u_i)_{i\in I}$ for E, by Proposition 1.10, for every $i \in I$, there is a unique linear form u_i^* such that

$$u_i^*(u_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for every $j \in I$. The linear form u_i^* is called the *coordinate form* of index i w.r.t. the basis $(u_i)_{i\in I}$.

Remark: Given an index set I, authors often define the so called *Kronecker symbol* $\delta_{i\,j}$, such that

$$\delta_{i\,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for all $i, j \in I$.

Then,

$$u_i^*(u_j) = \delta_{i\,j}.$$

The reason for the terminology *coordinate form* is as follows: If E has finite dimension and if $(u_1, \ldots, u_n)$ is a basis of E, for any vector

$$v = \lambda_1 u_1 + \cdots + \lambda_n u_n,$$

we have

$$u_i^*(v) = \lambda_i.$$

Therefore, u_i^* is the linear function that returns the ith coordinate of a vector expressed over the basis $(u_1, \ldots, u_n)$.

As part of the Duality theorem (Theorem 3.14), we will prove that if $(u_1, \ldots, u_n)$ is a basis of E, then $(u_1^*, \ldots, u_n^*)$ is a basis of E^* called the *dual basis*.

If $(u_1, \ldots, u_n)$ is a basis of $\mathbb{R}^n$ (more generally K^n), it is possible to find explicitly the dual basis $(u_1^*, \ldots, u_n^*)$, where each u_i^* is represented by a row vector.

For example, consider the columns of the Bézier matrix

$$B_4 = \begin{pmatrix} 1 & -3 & 3 & -1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The form u_1^* is represented by a row vector $(\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4)$ such that

$$\begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{pmatrix} \begin{pmatrix} 1 & -3 & 3 & -1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}.$$

This implies that u_1^* is the first row of the inverse of B_4.

Since

$$B_4^{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1/3 & 2/3 & 1 \\ 0 & 0 & 1/3 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

the linear forms $(u_1^*, u_2^*, u_3^*, u_4^*)$ correspond to the rows of B_4^{-1}.

In particular, u_1^* is represented by $(1\ 1\ 1\ 1)$.

The above method works for any n. Given any basis $(u_1, \ldots, u_n)$ of $\mathbb{R}^n$, if P is the $n \times n$ matrix whose jth column is u_j, then the dual form u_i^* is given by the ith row of the matrix P^{-1}.

We have the following important duality theorem adapted from E. Artin.

Theorem 3.14. *(Duality theorem) Let E be a vector space of dimension n. The following properties hold:*

(a) For every basis $(u_1, \ldots, u_n)$ of E, the family of coordinate forms $(u_1^*, \ldots, u_n^*)$ is a basis of E^*.

(b) For every subspace V of E, we have $V^{00} = V$.

(c) For every pair of subspaces V and W of E such that $E = V \oplus W$, with V of dimension m, for every basis $(u_1, \ldots, u_n)$ of E such that $(u_1, \ldots, u_m)$ is a basis of V and $(u_{m+1}, \ldots, u_n)$ is a basis of W, the family $(u_1^*, \ldots, u_m^*)$ is a basis of the orthogonal W^0 of W in E^*. Furthermore, we have $W^{00} = W$, and

$$\dim(W) + \dim(W^0) = \dim(E).$$

(d) For every subspace U of E^*, we have

$$\dim(U) + \dim(U^0) = \dim(E),$$

where U^0 is the orthogonal of U in E, and $U^{00} = U$.

Part (a) of Theorem 3.14 shows that

$$\dim(E) = \dim(E^*),$$

and if $(u_1, \ldots, u_n)$ is a basis of E, then $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space E^* called the *dual basis* of $(u_1, \ldots, u_n)$.

Define the function $\mathcal{E}$ from subspaces of E to subspaces of E^* and the function $\mathcal{Z}$ from subspaces of E^* to subspaces of E by

$$\begin{aligned} \mathcal{E}(V) &= V^0, \quad V \subseteq E \\ \mathcal{Z}(U) &= U^0, \quad U \subseteq E^*. \end{aligned}$$

By part (c) and (d) of theorem 3.14,

$$\begin{aligned} (\mathcal{Z} \circ \mathcal{E})(V) &= V^{00} = V \\ (\mathcal{E} \circ \mathcal{Z})(U) &= U^{00} = U, \end{aligned}$$

so $\mathcal{Z} \circ \mathcal{E} = \mathrm{id}$ and $\mathcal{E} \circ \mathcal{Z} = \mathrm{id}$, and the maps $\mathcal{E}$ and $\mathcal{V}$ are inverse bijections.

These maps set up a *duality* between subspaces of E, and subspaces of E^*.

One should be careful that this bijection does not hold if E has infinite dimension. Some restrictions on the dimensions of U and V are needed.

Suppose that V is a subspace of $\mathbb{R}^n$ of dimension m and that $(v_1, \ldots, v_m)$ is a basis of V.

To find a basis of V^0, we first extend $(v_1, \ldots, v_m)$ to a basis $(v_1, \ldots, v_n)$ of $\mathbb{R}^n$, and then by part (c) of Theorem 3.14, we know that $(v_{m+1}^*, \ldots, v_n^*)$ is a basis of V^0.

For example, suppose that V is the subspace of $\mathbb{R}^4$ spanned by the two linearly independent vectors

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix},$$

the first two vectors of the Haar basis in $\mathbb{R}^4$.

The four columns of the Haar matrix

$$W = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{pmatrix}$$

form a basis of $\mathbb{R}^4$, and the inverse of W is given by

$$W^{-1} = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & -1/4 & -1/4 \\ 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \end{pmatrix}.$$

Since the dual basis $(v_1^*, v_2^*, v_3^*, v_4^*)$ is given by the row of W^{-1}, the last two rows of W^{-1},

$$\begin{pmatrix} 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \end{pmatrix},$$

form a basis of V^0.

We also obtain a basis by rescaling by the factor $1/2$, so the linear forms given by the row vectors

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

form a basis of V^0, the space of linear forms (linear equations) that vanish on the subspace V.

The method that we described to find V^0 requires first extending a basis of V and then inverting a matrix, but there is a more direct method.

Indeed, let A be the $n \times m$ matrix whose columns are the basis vectors $(v_1, \ldots, v_m)$ of V. Then, a linear form u represented by a row vector belongs to V^0 iff $uv_i = 0$ for $i = 1, \ldots, m$ iff

$$uA = 0$$

iff

$$A^\top u^\top = 0.$$

Therefore, all we need to do is to *find a basis of the nullspace of* $A^\top$.

This can be done quite effectively using the reduction of a matrix to reduced row echelon form (rref); see Section 4.5.

Here is another example illustrating the power of Theorem 3.14.

Let $E = \mathrm{M}_n(\mathbb{R})$, and consider the equations asserting that the sum of the entries in every row of a matrix $\in \mathrm{M}_n(\mathbb{R})$ is equal to the same number.

We have $n-1$ equations

$$\sum_{j=1}^{n}(a_{ij} - a_{i+1j}) = 0, \quad 1 \leq i \leq n-1,$$

and it is easy to see that they are linearly independent.

Therefore, the space U of linear forms in E^* spanned by the above linear forms (equations) has dimension $n-1$, and the space U^0 of matrices sastisfying all these equations has dimension $n^2 - n + 1$.

It is not so obvious to find a basis for this space.

When E is of finite dimension n and $(u_1, \ldots, u_n)$ is a basis of E, we noted that the family $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space E^*,

Let us see how the coordinates of a linear form $\varphi^* \in E^*$ over the basis $(u_1^*, \ldots, u_n^*)$ vary under a change of basis.

Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two bases of E, and let $P = (a_{i\,j})$ be the change of basis matrix from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$, so that

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i.$$

If

$$\varphi^* = \sum_{i=1}^{n} \varphi_i u_i^* = \sum_{i=1}^{n} \varphi_i' v_i^*,$$

after some algebra, we get

$$\varphi_j' = \sum_{i=1}^{n} a_{i\,j} \varphi_i.$$

Comparing with the change of basis

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i,$$

we note that this time, the coordinates (φ_i) of the linear form φ^* change in the *same direction* as the change of basis.

For this reason, we say that the coordinates of linear forms are *covariant*.

By abuse of language, it is often said that linear forms are *covariant*, which explains why the term *covector* is also used for a linear form.

Observe that if $(e_1, \ldots, e_n)$ is a basis of the vector space E, then, as a linear map from E to K, every linear form $f \in E^*$ is represented by a $1 \times n$ matrix, that is, by a *row vector*

$$(\lambda_1 \cdots \lambda_n),$$

with respect to the basis $(e_1, \ldots, e_n)$ of E, and 1 of K, where $f(e_i) = \lambda_i$.

A vector $u = \sum_{i=1}^{n} u_i e_i \in E$ is represented by a $n \times 1$ matrix, that is, by a *column vector*

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

and the action of f on u, namely $f(u)$, is represented by the matrix product

$$\begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \lambda_1 u_1 + \cdots + \lambda_n u_n.$$

On the other hand, with respect to the dual basis $(e_1^*, \ldots, e_n^*)$ of E^*, the linear form f is represented by the column vector

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

We will now pin down the relationship between a vector space E and its bidual E^{**}.

Proposition 3.15. *Let E be a vector space. The following properties hold:*

(a) The linear map $\mathrm{eval}_E \colon E \to E^{**}$ *defined such that*

$$\mathrm{eval}_E(v) = \mathrm{eval}_v, \quad \textit{for all } v \in E,$$

that is, $\mathrm{eval}_E(v)(u^*) = \langle u^*, v\rangle = u^*(v)$ *for every* $u^* \in E^*$*, is injective.*

(b) When E is of finite dimension n, the linear map $\mathrm{eval}_E \colon E \to E^{**}$ *is an isomorphism (called the canonical isomorphism).*

When E is of finite dimension and $(u_1, \ldots, u_n)$ is a basis of E, in view of the canonical isomorphism
$\mathrm{eval}_E \colon E \to E^{**}$, the basis $(u_1^{**}, \ldots, u_n^{**})$ of the bidual is identified with $(u_1, \ldots, u_n)$.

Proposition 3.15 can be reformulated very fruitfully in terms of pairings.

Definition 3.8. Given two vector spaces E and F over K, a *pairing between E and F* is a bilinear map $\varphi\colon E \times F \to K$. Such a pairing is *nondegenerate* iff

(1) for every $u \in E$, if $\varphi(u, v) = 0$ for all $v \in F$, then $u = 0$, and

(2) for every $v \in F$, if $\varphi(u, v) = 0$ for all $u \in E$, then $v = 0$.

A pairing $\varphi\colon E \times F \to K$ is often denoted by $\langle -, - \rangle\colon E \times F \to K$.

For example, the map $\langle -, - \rangle\colon E^* \times E \to K$ defined earlier is a nondegenerate pairing (use the proof of (a) in Proposition 3.15).

Given a pairing $\varphi\colon E \times F \to K$, we can define two maps $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ as follows:

For every $u \in E$, we define the linear form $l_\varphi(u)$ in F^* such that

$$l_\varphi(u)(y) = \varphi(u, y) \quad \text{for every } y \in F,$$

and for every $v \in F$, we define the linear form $r_\varphi(v)$ in E^* such that

$$r_\varphi(v)(x) = \varphi(x, v) \quad \text{for every } x \in E.$$

Proposition 3.16. *Given two vector spaces E and F over K, for every nondegenerate pairing $\varphi\colon E \times F \to K$ between E and F, the maps $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ are linear and injective. Furthermore, if E and F have finite dimension, then this dimension is the same and $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ are bijections.*

When E has finite dimension, the nondegenerate pairing $\langle -, - \rangle\colon E^* \times E \to K$ yields another proof of the existence of a natural isomorphism between E and E^{**}.

Interesting nondegenerate pairings arise in exterior algebra.

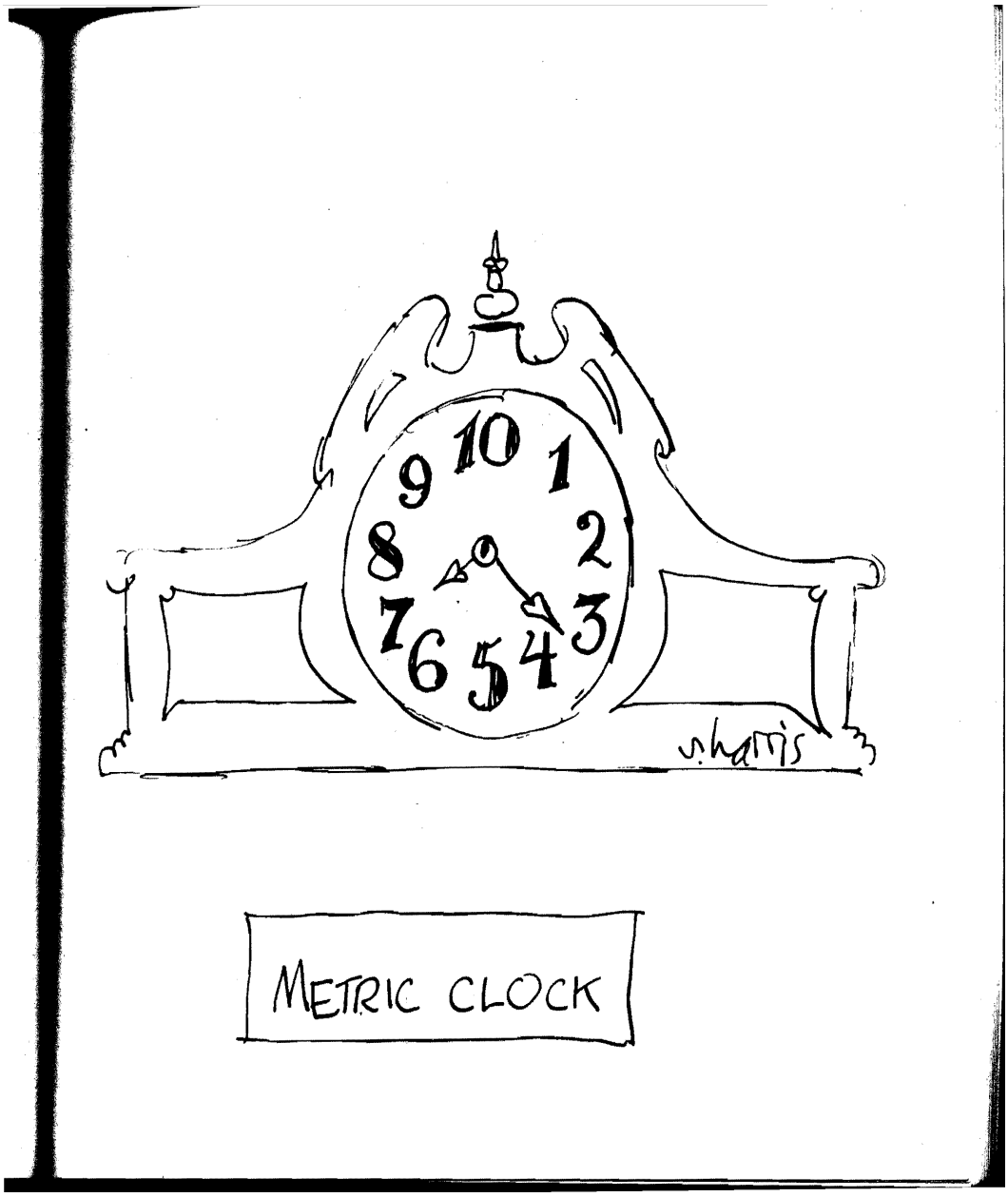

Figure 3.4: Metric Clock

3.4 Hyperplanes and Linear Forms

Actually, Proposition 3.17 below follows from parts (c) and (d) of Theorem 3.14, but we feel that it is also interesting to give a more direct proof.

Proposition 3.17. *Let E be a vector space. The following properties hold:*

(a) Given any nonnull linear form $f^* \in E^*$, its kernel $H = \mathrm{Ker}\, f^*$ is a hyperplane.

(b) For any hyperplane H in E, there is a (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$.

(c) Given any hyperplane H in E and any (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$, for every linear form $g^* \in E^*$, $H = \mathrm{Ker}\, g^*$ iff $g^* = \lambda f^*$ for some $\lambda \neq 0$ in K.

We leave as an exercise the fact that every subspace $V \neq E$ of a vector space E, is the intersection of all hyperplanes that contain V.

We now consider the notion of transpose of a linear map and of a matrix.

3.5 Transpose of a Linear Map and of a Matrix

Given a linear map $f\colon E \to F$, it is possible to define a map $f^\top\colon F^* \to E^*$ which has some interesting properties.

Definition 3.9. Given a linear map $f\colon E \to F$, the *transpose* $f^\top\colon F^* \to E^*$ of f is the linear map defined such that

$$f^\top(v^*) = v^* \circ f,$$

for every $v^* \in F^*$.

Equivalently, the linear map $f^\top\colon F^* \to E^*$ is defined such that

$$\langle v^*, f(u)\rangle = \langle f^\top(v^*), u\rangle,$$

for all $u \in E$ and all $v^* \in F^*$.

It is easy to verify that the following properties hold:

$$\begin{aligned}(f+g)^\top &= f^\top + g^\top \\ (g \circ f)^\top &= f^\top \circ g^\top \\ \mathrm{id}_E^\top &= \mathrm{id}_{E^*}.\end{aligned}$$

Note the reversal of composition on the right-hand side of $(g \circ f)^\top = f^\top \circ g^\top$.

The equation $(g \circ f)^\top = f^\top \circ g^\top$ implies the following useful proposition.

Proposition 3.18. *If $f\colon E \to F$ is any linear map, then the following properties hold:*

(1) If f is injective, then $f^\top$ is surjective.

(2) If f is surjective, then $f^\top$ is injective.

We also have the following property showing the naturality of the eval map.

Proposition 3.19. *For any linear map $f\colon E \to F$, we have*

$$f^{\top\top} \circ \mathrm{eval}_E = \mathrm{eval}_F \circ f,$$

or equivalently, the following diagram commutes:

$$\begin{tikzcd}
E^{**} \arrow[r, "f^{\top\top}"] & F^{**} \\
E \arrow[u, "\mathrm{eval}_E"] \arrow[r, "f"'] & F. \arrow[u, "\mathrm{eval}_F"']
\end{tikzcd}$$

If E and F are finite-dimensional, then eval_E and eval_F are isomorphisms, so Proposition 3.19 shows that

$$f^{\top\top} = \mathrm{eval}_F^{-1} \circ f \circ \mathrm{eval}_E. \tag{$*$}$$

The above equation is often interpreted as follows: if we identify E with its bidual E^{**} and F with its bidual F^{**}, then $f^{\top\top} = f$.

This is an abuse of notation; the rigorous statement is $(*)$.

Proposition 3.20. *Given a linear map $f\colon E \to F$, for any subspace V of E, we have*

$$f(V)^0 = (f^\top)^{-1}(V^0) = \{w^* \in F^* \mid f^\top(w^*) \in V^0\}.$$

As a consequence,

$$\mathrm{Ker}\, f^\top = (\mathrm{Im}\, f)^0 \quad \textit{and} \quad \mathrm{Ker}\, f = (\mathrm{Im}\, f^\top)^0.$$

The following theorem shows the relationship between the rank of f and the rank of $f^\top$.

Theorem 3.21. *Given a linear map $f\colon E \to F$, the following properties hold.*

(a) The dual $(\mathrm{Im}\, f)^*$ of $\mathrm{Im}\, f$ is isomorphic to $\mathrm{Im}\, f^\top = f^\top(F^*)$; that is,

$$(\mathrm{Im}\, f)^* \approx \mathrm{Im}\, f^\top.$$

(b) If F is finite dimensional, then $\mathrm{rk}(f) = \mathrm{rk}(f^\top)$.

The following proposition can be shown, but it requires a generalization of the duality theorem.

Proposition 3.22. *If $f\colon E \to F$ is any linear map, then the following identities hold:*

$$\begin{aligned}
\mathrm{Im}\, f^\top &= (\mathrm{Ker}\,(f))^0 \\
\mathrm{Ker}\,(f^\top) &= (\mathrm{Im}\, f)^0 \\
\mathrm{Im}\, f &= (\mathrm{Ker}\,(f^\top)^0 \\
\mathrm{Ker}\,(f) &= (\mathrm{Im}\, f^\top)^0.
\end{aligned}$$

The following proposition shows the relationship between the matrix representing a linear map $f\colon E \to F$ and the matrix representing its transpose $f^\top\colon F^* \to E^*$.

Proposition 3.23. *Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for E, and $(v_1, \ldots, v_m)$ be a basis for F. Given any linear map $f\colon E \to F$, if $M(f)$ is the $m \times n$-matrix representing f w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, the $n \times m$-matrix $M(f^\top)$ representing $f^\top\colon F^* \to E^*$ w.r.t. the dual bases $(v_1^*, \ldots, v_m^*)$ and $(u_1^*, \ldots, u_n^*)$ is the transpose $M(f)^\top$ of $M(f)$.*

We now can give a very short proof of the fact that the rank of a matrix is equal to the rank of its transpose.

Proposition 3.24. *Given a $m \times n$ matrix A over a field K, we have* $\mathrm{rk}(A) = \mathrm{rk}(A^\top)$.

Thus, given an $m \times n$-matrix A, the maximum number of linearly independent columns is equal to the maximum number of linearly independent rows.

Proposition 3.25. *Given any $m \times n$ matrix A over a field K (typically $K = \mathbb{R}$ or $K = \mathbb{C}$), the rank of A is the maximum natural number r such that there is an invertible $r \times r$ submatrix of A obtained by selecting r rows and r columns of A.*

For example, the 3×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

has rank 2 iff one of the three 2×2 matrices

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix} \quad \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

is invertible. We will see in Chapter 5 that this is equivalent to the fact the determinant of one of the above matrices is nonzero.

This is not a very efficient way of finding the rank of a matrix. We will see that there are better ways using various decompositions such as LU, QR, or SVD.

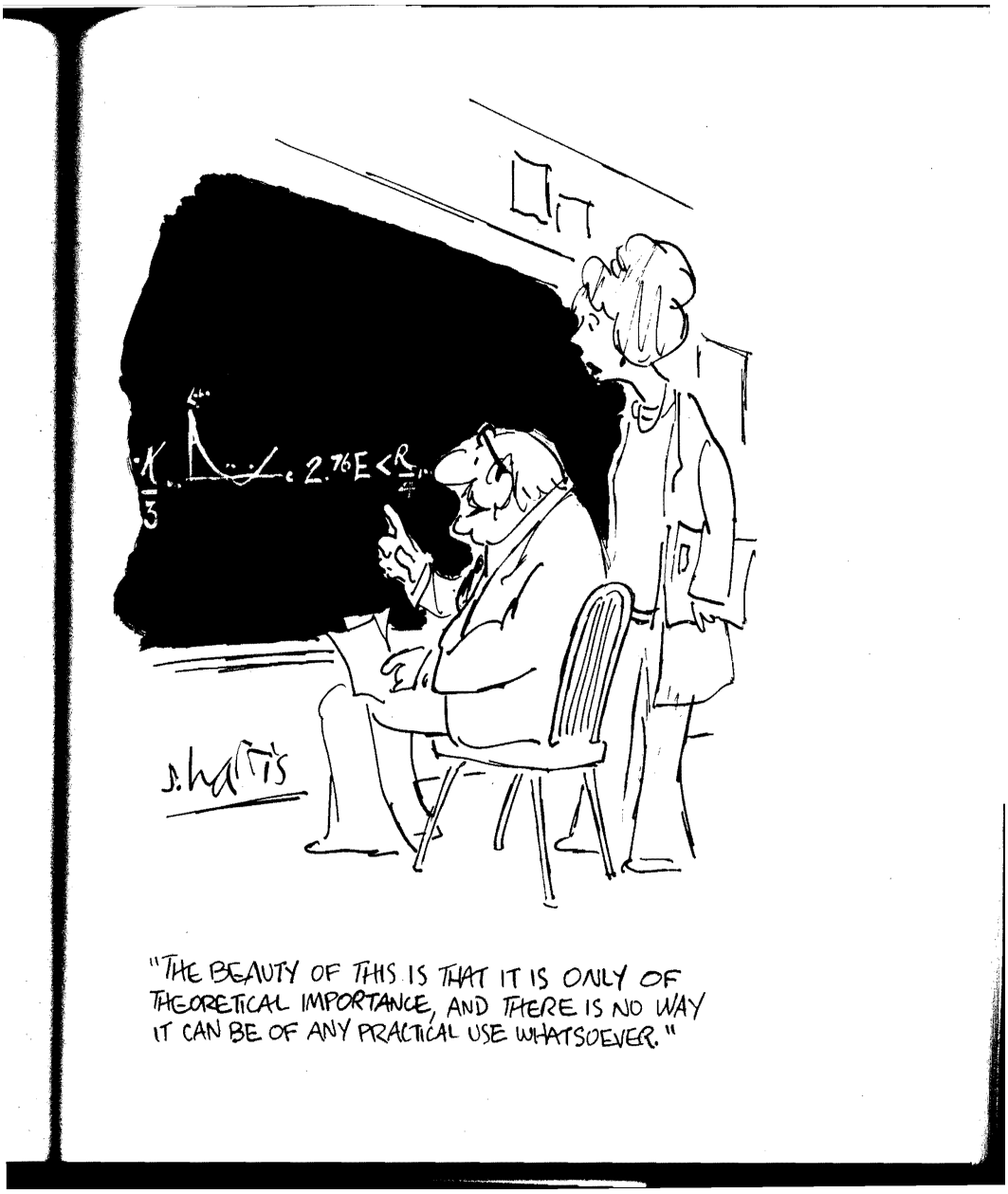

Figure 3.5: Beauty

3.6 The Four Fundamental Subspaces

Given a linear map $f\colon E \to F$ (where E and F are finite-dimensional), Proposition 3.20 revealed that the four spaces

$$\mathrm{Im}\, f,\ \mathrm{Im}\, f^{\top},\ \mathrm{Ker}\, f,\ \mathrm{Ker}\, f^{\top}$$

play a special role. They are often called the *fundamental subspaces* associated with f.

These spaces are related in an intimate manner, since Proposition 3.20 shows that

$$\begin{aligned} \mathrm{Ker}\, f &= (\mathrm{Im}\, f^{\top})^0 \\ \mathrm{Ker}\, f^{\top} &= (\mathrm{Im}\, f)^0, \end{aligned}$$

and Theorem 3.21 shows that

$$\mathrm{rk}(f) = \mathrm{rk}(f^{\top}).$$

It is instructive to translate these relations in terms of matrices (actually, certain linear algebra books make a big deal about this!).

If $\dim(E) = n$ and $\dim(F) = m$, given any basis $(u_1, \ldots, u_n)$ of E and a basis $(v_1, \ldots, v_m)$ of F, we know that f is represented by an $m \times n$ matrix $A = (a_{i\,j})$, where the jth column of A is equal to $f(u_j)$ over the basis $(v_1, \ldots, v_m)$.

Furthermore, the transpose map $f^\top$ is represented by the $n \times m$ matrix $A^\top$ (with respect to the dual bases).

Consequently, the four fundamental spaces

$$\mathrm{Im}\, f,\ \mathrm{Im}\, f^\top,\ \mathrm{Ker}\, f,\ \mathrm{Ker}\, f^\top$$

correspond to

(1) The *column space* of A, denoted by $\mathrm{Im}\, A$ or $\mathcal{R}(A)$; this is the subspace of $\mathbb{R}^m$ spanned by the columns of A, which corresponds to the image $\mathrm{Im}\, f$ of f.

(2) The *kernel* or *nullspace* of A, denoted by $\mathrm{Ker}\, A$ or $\mathcal{N}(A)$; this is the subspace of $\mathbb{R}^n$ consisting of all vectors $x \in \mathbb{R}^n$ such that $Ax = 0$.

(3) The *row space* of A, denoted by $\mathrm{Im}\, A^\top$ or $\mathcal{R}(A^\top)$; this is the subspace of $\mathbb{R}^n$ spanned by the rows of A, or equivalently, spanned by the columns of $A^\top$, which corresponds to the image $\mathrm{Im}\, f^\top$ of $f^\top$.

(4) The *left kernel* or *left nullspace* of A denoted by $\mathrm{Ker}\, A^\top$ or $\mathcal{N}(A^\top)$; this is the kernel (nullspace) of $A^\top$, the subspace of $\mathbb{R}^m$ consisting of all vectors $y \in \mathbb{R}^m$ such that $A^\top y = 0$, or equivalently, $y^\top A = 0$.

Recall that the dimension r of $\mathrm{Im}\, f$, which is also equal to the dimension of the column space $\mathrm{Im}\, A = \mathcal{R}(A)$, is the *rank* of A (and f).

Then, some our previous results can be reformulated as follows:

1. The column space $\mathcal{R}(A)$ of A has dimension r.
2. The nullspace $\mathcal{N}(A)$ of A has dimension $n - r$.
3. The row space $\mathcal{R}(A^\top)$ has dimension r.
4. The left nullspace $\mathcal{N}(A^\top)$ of A has dimension $m - r$.

The above statements constitute what Strang calls the *Fundamental Theorem of Linear Algebra, Part I* (see Strang [32]).

The two statements

$$\begin{aligned} \mathrm{Ker}\, f &= (\mathrm{Im}\, f^{\top})^0 \\ \mathrm{Ker}\, f^{\top} &= (\mathrm{Im}\, f)^0 \end{aligned}$$

translate to

(1) The nullspace of A is the orthogonal of the row space of A.

(2) The left nullspace of A is the orthogonal of the column space of A.

The above statements constitute what Strang calls the *Fundamental Theorem of Linear Algebra, Part II* (see Strang [32]).

Since vectors are represented by column vectors and linear forms by row vectors (over a basis in E or F), a vector $x \in \mathbb{R}^n$ is orthogonal to a linear form y if

$$yx = 0.$$

Then, a vector $x \in \mathbb{R}^n$ is orthogonal to the row space of A iff x is orthogonal to every row of A, namely $Ax = 0$, which is equivalent to the fact that x belong to the nullspace of A.

Similarly, the column vector $y \in \mathbb{R}^m$ (representing a linear form over the dual basis of F^*) belongs to the nullspace of $A^\top$ iff $A^\top y = 0$, iff $y^\top A = 0$, which means that the linear form given by $y^\top$ (over the basis in F) is orthogonal to the column space of A.

Since (2) is equivalent to the fact that *the column space of A is equal to the orthogonal of the left nullspace of A*, we get the following criterion for the solvability of an equation of the form $Ax = b$:

The equation $Ax = b$ has a solution iff for all $y \in \mathbb{R}^m$, if $A^\top y = 0$, then $y^\top b = 0$.

Indeed, the condition on the right-hand side says that b is orthogonal to the left nullspace of A, that is, that b belongs to the column space of A.

This criterion can be cheaper to check that checking directly that b is spanned by the columns of A. For example, if we consider the system

$$\begin{aligned} x_1 - x_2 &= b_1 \\ x_2 - x_3 &= b_2 \\ x_3 - x_1 &= b_3 \end{aligned}$$

which, in matrix form, is written $Ax = b$ as below:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

we see that the rows of the matrix A add up to 0.

In fact, it is easy to convince ourselves that the left nullspace of A is spanned by $y = (1, 1, 1)$, and so the system is solvable iff $y^\top b = 0$, namely

$$b_1 + b_2 + b_3 = 0.$$

Note that the above criterion can also be stated negatively as follows:

The equation $Ax = b$ has no solution iff there is some $y \in \mathbb{R}^m$ such that $A^\top y = 0$ and $y^\top b \neq 0$.

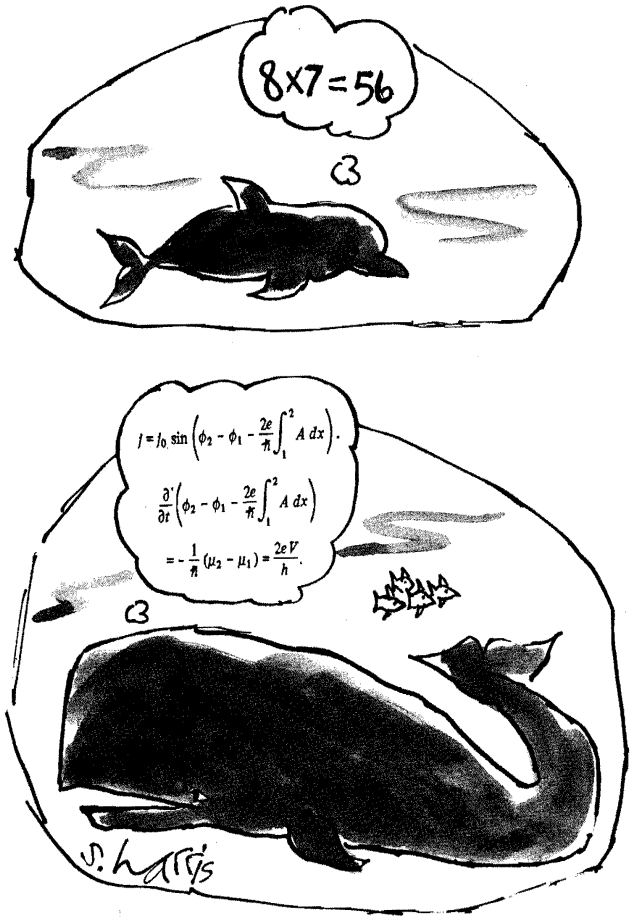

Figure 3.6: Brain Size?