2 Linear Spaces and Transformations

2.1 Vector Spaces

We showed in Sect. 1.3 that when $\mathcal{F}$ was a field, the n-fold product set $\mathcal{F}^n$ had an additional operation defined on it, which was induced from addition in $\mathcal{F}$, so that $(\mathcal{F}^n, +)$ became an abelian group with zero $\underline{0}$. Moreover we were able to define a product $\cdot : \mathcal{F} \times \mathcal{F}^n \to \mathcal{F}^n$ which takes (α, x) to a new element of $\mathcal{F}^n$ called (αx). Elements of $\mathcal{F}^n$ are known as *vectors*, and elements of $\mathcal{F}$ as *scalars*. The properties that we discovered in $\mathcal{F}^n$ characterise a *vector space*. A vector space is also known as a linear space.

Definition 2.1 A *vector space* $(V, +)$ is an abelian additive group with zero $\underline{0}$, together with a field $(\mathcal{F}, +, \cdot)$ with zero 0 and identity 1. An element of V is called a *vector* and an element of $\mathcal{F}$ a *scalar*. Moreover for any $\alpha \in \mathcal{F}, v \in V$ there is a scalar multiplication $(\alpha, v) \to \alpha v \in V$ which satisfies the following properties:

V1. $\alpha(v_1 + v_2) = \alpha v_1 + \alpha v_2$, for any $\alpha \in \mathcal{F}, v_1, v_2 \in V$.

V2. $(\alpha + \beta)v = \alpha v + \beta v$, for any $\alpha, \beta \in \mathcal{F}, v \in V$.

V3. $(\alpha\beta)v = \alpha(\beta v)$, for any $\alpha, \beta \in \mathcal{F}, v \in V$.

V4. $1 \cdot v = v$, for $1 \in \mathcal{F}$, and for any $v \in V$.

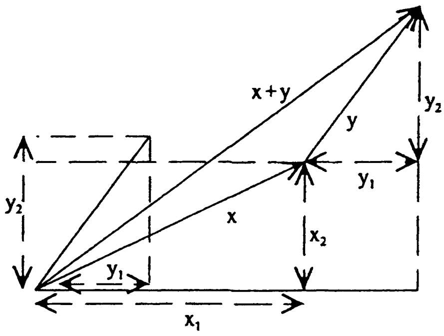
Call V a vector space over the field $\mathcal{F}$. From the previous discussion the set $\Re^n$ becomes an abelian group $(\Re^n, +)$ under addition. We shall frequently

write

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

N. Schofield, *Mathematical Methods in Economics and Social Choice*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-39818-6_2,

Fig. 2.1 Vector addition



for a vector in $\Re^n$, where $x_1, \ldots, x_n$ are called the coordinates of x. Vector addition as in Fig. 2.1 is then defined by

$$x + y = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

A vector space over $\Re$ is called a *real vector space*.

For example $(\mathcal{Z}_2, +, \cdot)$ is a field and so $(\mathcal{Z}_2)^n$ is a vector space over the field $\mathcal{Z}_2$. It may not be possible to represent each vector in a vector space by a list of coordinates. For example, consider the set of all functions with domain X and image in $\Re$. Call this set $\Re^X$. If $f, g \in \Re^X$, define $f + g$ to be that function which maps $x \in X$ to $f(x) + g(x)$. Clearly there is a zero function $\underline{0}$ defined by $\underline{0}(x) = \underline{0}$, and each f has an inverse $(-f)$ defined by $(-f)(x) = -(f(x))$. Finally for $\alpha \in \Re$, $f \in \Re^X$, define $\alpha f : X \to \Re$ by $(\alpha f)(x) = \alpha(f(x))$. Thus $\Re^X$ is a vector space over $\Re$.

Definition 2.2 Let $(V, +)$ be a vector space over a field, $\mathcal{F}$. A subset V' of V is called a *vector subspace* of V if and only if
1. $v_1, v_2 \in V' \Rightarrow v_1 + v_2 \in V'$, and
2. if $\alpha \in F$ and $v \in V'$ then $\alpha v' \in V'$.

Lemma 2.1 *If $(V, +)$ is a vector space with zero $\underline{0}$ and V' is a vector subspace, then, for each $v \in V'$, the inverse $(-v) \in V'$, and $\underline{0} \in V'$, so $(V', +)$ is a subgroup of $(V, +)$.*

Proof Suppose $v \in V'$. Since $\mathcal{F}$ is a field, there is an identity 1, with additive inverse -1. But by **V2**, $(1-1)v = 1 \cdot v + (-1)v = 0 \cdot v$, since $1 - 1 = 0$. Now $(1+0)v = 1 \cdot v + 0 \cdot v$, and so $0 \cdot v = \underline{0}$. Thus $(-1)v = (-v)$. Since V' is a vector subspace, $(-1)v \in V'$, and so $(-v) \in V'$. But then $v + (-v) = \underline{0}$, and so $\underline{0} \in V'$. $\square$

From now on we shall simply write V for a vector space, and refer to the field only on occasion.

Definition 2.3 Let $V' = \{v_1, \ldots, v_r\}$ be a set of vectors in the vector space V. A vector v is called a *linear combination* of the set V' iff v can be written in the form

$$v = \sum_{i=1}^{r} \lambda_i v_i$$

where each $\lambda_i, i = 1, \ldots, r$ belongs to the field $\mathcal{F}$. The *span* of V', written $\mathrm{Span}(V')$ is the set of vectors which are linear combinations of the set V'. If $V'' = \mathrm{Span}(V')$, then V' is said *to span* V''.

For example, suppose

$$V' = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}.$$

Since we can solve the equation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

for any $(x, y) \in \Re^2$, by setting $\alpha = \frac{1}{3}(2y - x)$ and $\beta = \frac{1}{3}(2x - y)$, it is clear that V' is a span for $\Re^2$.

Lemma 2.2 *If V' is a finite set of vectors in the vector space, V, then* $\mathrm{Span}(V')$ *is a vector subspace of V.*

Proof We seek to show that for any $\alpha, \beta \in \mathcal{F}$ and any $u, w \in \mathrm{Span}(V')$, then $\alpha u + \beta w \in \mathrm{Span}(V')$. By definition, if $V' = \{v_1, \ldots, v_r\}$, then $u = \sum_{i=1}^{r} \eta_i v_i$ and $w = \sum_{i=1}^{r} \mu_i v_i$, where $\eta_i, \mu_i \in \mathcal{F}$ for $i = 1, \ldots, r$. But then $\alpha u + \beta w = \alpha \sum_{i=1}^{r} \eta_i v_i + \beta \sum_{i=1}^{r} \mu_i v_i = \sum_{i=1}^{r} \lambda_i v_i$, where $\lambda_i = \alpha\eta_i + \beta\mu_i \in \mathcal{F}$, for $i = 1, \ldots, r$. Thus $\alpha u + \beta w \in \mathrm{Span}(V')$.

Note that, by this lemma, the zero vector $\underline{0}$ belongs to $\mathrm{Span}(V')$. □

Definition 2.4 Let $V' = \{v_1, \ldots, v_r\}$ be a set of vectors in V. V' is called a *frame* iff $\sum_{i=1}^{r} \alpha_i v_i = \underline{0}$ implies that $\alpha_i = 0$ for $i = 1, \ldots, r$. (Here each α_i belongs to the field $\mathcal{F}$.) In this case the set V' is called a *linearly independent set*. If V' is not a frame, the vectors in V' are said to be linearly dependent. Say a vector is *linearly dependent* on $V' = \{v_1, \ldots, v_r\}$ iff $v \in \mathrm{Span}(V')$.

Note that if V' is a frame, then

1. $\underline{0} \notin V'$ since $\alpha\underline{0} = \underline{0}$ for every non-zero $\alpha \in F$.
2. If $v \in V'$ then $(-v) \notin V'$, otherwise $1 \cdot v + 1(-v) = \underline{0}$ would belong to V', contradicting (1).

Lemma 2.3
1. *V' is not a frame iff there is some vector $v \in V'$ which is linearly dependent on $V' \backslash \{v\}$.*
2. *If V' is a frame, then any subset of V' is a frame.*
3. *If V' spans V'', but V' is not a frame, then there exists some vector $v \in V'$ such that $V''' = V' \backslash \{v\}$ spans V''.*

Proof Let $V' = \{v_1, \ldots, v_r\}$ be the set of vectors in the vector space V.
1. Suppose V' is not a frame. Then there exists an equation $\sum_{j=1}^{r} \alpha_j v_j = 0$, where, for at least one k, it is the case that $\alpha_k \neq 0$. But then $v_k = -\frac{1}{\alpha_k}(\sum_{j \neq k} \alpha_j v_j)$. Let $v_k = v$. Then v is linearly dependent on $V' \backslash \{v\}$. On the other hand suppose that v_1, say, is linearly dependent on $\{v_2, \ldots, v_r\}$. Then $v_1 = \sum_{j=2}^{r} \alpha_j v_j$, and so $\underline{0} = -v_1 + \sum_{j=2}^{r} \alpha_j v_j = \sum_{j=1}^{r} \alpha_j v_j$ where $\alpha_1 = -1$. Since $\alpha_1 \neq 0$, V' cannot be a frame.
2. Suppose V'' is a subset of V', but that V'' is not a frame. For convenience let $V'' = \{v_1, \ldots, v_k\}$ where $k \leq r$. Then there is a non-zero solution
$$\underline{0} \neq \sum_{j=1}^{k} \alpha_j v_j.$$
Since V'' is a subset of V', this implies that V' cannot be a frame. Thus if V' is a frame, so is any subset V''.
3. Suppose that V' is not a frame, but that it spans V''. By part (1), there exists a vector v_1, say, in V' such that v_1 belongs to $\mathrm{Span}(V' \backslash \{v_1\})$. Thus $v_1 = \sum_{j=2}^{r} \alpha_j v_j$. Since V' is a span for V'', any vector v in V'' can be written
$$\begin{aligned} v &= \sum_{j=1}^{r} \beta_j v_j \\ &= \beta_1 \left(\sum_{j=2}^{r} \alpha_j v_j \right) + \sum_{j=2}^{r} \beta_j v_j. \end{aligned}$$
Thus v is a linear combination of $V' \backslash \{v_1\}$ and so $V'' = \mathrm{Span}(V' \backslash \{v_1\})$.

Let $V''' = V' \backslash \{v_1\}$ to complete the proof. □

Definition 2.5 A *basis* for a vector space V is a frame V' which spans V.

For example, we previously considered
$$V' = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

and showed that any vector in $\Re^2$ could be written as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \left(\frac{2y-x}{3}\right)\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \left(\frac{2x-y}{3}\right)\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Thus V' is a span for $\Re^2$. Moreover if $(x, y) = (0, 0)$ then $\lambda_1 = \lambda_2 = 0$ and so V' is a frame. Hence V' is a basis for $\Re^2$. If $V' = \{v_1, \dots, v_n\}$ is a basis for a vector space V then any vector $v \in V$ can be written

$$v = \sum_{j=1}^{n} \alpha_j v_j$$

and the elements $(\alpha_1, \dots, \alpha_n)$ are known as the *coordinates* of the vector v, *with respect to* the basis V'.

For example the *natural basis* for $\Re^n$ is the set $V' = \{e_1, \dots, e_n\}$ where $e_i = (0, \dots, 1, \dots, 0\}$ with a 1 in the ith position.

Lemma 2.4 *$\{e_1, \dots, e_n\}$ is a basis for $\Re^n$.*

Proof We can write any vector x in $\Re^n$ as $\{x_1, \dots, x_n\}$. Clearly

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \cdot \end{pmatrix} + \cdots x_n \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}.$$

If $x = 0$ then $x_1 = \cdots = x_n = 0$ and so $\{e_1, \dots, e_n\}$ is a frame, as well as a span, and thus a basis for $\Re^n$. □

However a single vector x will have different coordinates depending on the basis chosen. For example the vector (x, y) has coordinates (x, y) in the basis $\{e_1, e_2\}$ but coordinates $(\frac{2y-x}{3}, \frac{2x-y}{3})$ with respect to the basis $\left\{\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)\right\}$.

Once the basis is chosen, the coordinates of any vector with respect to that basis are unique. □

Lemma 2.5 *Suppose $V' = \{v_1, \dots, v_n\}$ is a basis for V. Let $v = \sum_{i=1}^n \alpha_i v_i$. Then the coordinates $(\alpha_1, \dots, \alpha_n)$, with respect to the basis, are unique.*

Proof If the coordinates were not unique then it would be possible to write $v = \sum_{i=1}^n \beta_i v_i = \sum_{i=1}^n \alpha_i v_i$ with $\beta_i \neq \alpha_i$ for some i.

But $\underline{0} = v - v = \sum_{i=1}^n \alpha_i v_i - \sum_{i=1}^n \beta_i v_i = \sum_{i=1}^n (\alpha_i - \beta_i) v_i$.

Since V' is a frame, $\alpha_i - \beta_i = 0$ for $i = 1, \dots, n$. Thus $\alpha_i = \beta_i$ for all i, and so the coordinates are unique. □

Note in particular that with respect to *any* basis $\{v_1, \dots, v_n\}$ for V, the unique zero vector $\underline{0}$ always has coordinates $(0, \dots, 0)$.

Definition 2.6 A space V is *finitely generated* iff there exists a span V', for V, which has a finite number of elements.

Lemma 2.6 *If V is a finitely generated vector space, then it has a basis with a finite number of elements.*

Proof Since V is finitely generated, there is a finite set $V_1 = \{v_1, \dots, v_n\}$ which spans V. If V_1 is a frame, then it is a basis. If V_1 is linearly dependent, then by Lemma 2.3(3) there is a vector $v \in V_1$, such that $\text{Span}(V_2) = V$, where $V_2 = V_1 \backslash \{v\}$. Again if V_2 is a frame, then it is a basis. If there were no subset $V_r = \{v_1, \dots, v_{n-r+1}\}$ of V_1 which was a frame, then V_1 would have to be the empty set, implying that V was an empty set. But this contradicts $\underline{0} \in V$. □

Lemma 2.7 *If V is a finitely generated vector space, and V_1 is a frame, then there is a basis V_2 for V which includes V_1.*

Proof Let $V_1 = \{v_1, \dots, v_r\}$. If $\text{Span}(V_1) = V$ then V_1 is a basis. So suppose that $\text{Span}(V_1) \neq V$. Then there exists an element $v_{r+1} \in V$ which does not belong to $\text{Span}(V_1)$. We seek to show that $V_2 = V_1 \cup \{v_{r+1}\}$ is a frame. Consider $\underline{0} = \alpha_{r+1} v_{r+1} + \sum_{i=1}^{r} \alpha_i v_i$.

If $\alpha_{r+1} = 0$, then the linear independence of V_1 implies that $\alpha_i = 0$, for $i = 1, \dots, r$. Thus V_2 is a frame. If $\alpha_{r+1} \neq 0$, then

$$v_{r+1} = -\frac{1}{\alpha_{r+1}} \left(\sum_{i=1}^{r} \alpha_i v_i \right).$$

But this implies that v_{r+1} belongs to $\text{Span}(V_1)$ and therefore that $V = \text{Span}(V_1)$. Thus V_2 is a frame. If V_2 is a span for V, then it is a basis. If V_2 is not a span, reiterate this process. Since V is finitely generated, there must be some frame $V_{n-r+1} = \{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ which is a span, and thus a basis for V. □

These two lemmas show that if V is a finitely generated vector space, and $\{v_1, \dots, v_m\}$ is a span then some subset $\{v_1, \dots, v_n\}$, with $n \leq m$, is a basis. A basis is a *minimal* span.

On the other hand if $X = \{v_1, \dots, v_r\}$ is a frame, but not a span, then elements may be added to X in such a way as to preserve linear independence, until this "superset" of X becomes a basis. Consequently a basis is a *maximal* frame. These two results can be combined into one theorem.

Exchange Theorem *Suppose that V is a finitely generated vector space. Let $X = \{x_1, \dots, x_n\}$ be a frame and $Y = \{y_1, \dots, y_n\}$ a span. Then there is some subset Y' of Y such that $X \cup Y'$ is a basis for V.*

Proof By induction, let $X_S = \{x_1, \dots, x_s\}$, for each $s = 1, \dots, n$, and let $X_0 = \phi$. □

We know already from Lemma 2.6 that there is some subset Y_0 of Y such that $X_0 \cup Y_0$ is a basis for V. Suppose for some $s < m$, there is a subset Y_s of Y such that $X_s \cup Y_s$ is a basis.

Let $Y_s = \{y_1, \dots, y_t\}$. Now $x_{s+1} \notin \mathrm{Span}(X_s \cup Y_s)$ since $X_s \cup Y_s$ is a basis. Thus $x_{s+1} = \sum_1^s \alpha_1 x_1 + \sum_1^t \beta_i y_i$. But $X_{s+1} = \{x_1, \dots, x_{s+1}\}$ is a frame, since it is a subset of X.

Thus at least one $\beta_j \neq 0$. Let $Y_{s+1} = Y_s \backslash \{y_j\}$, so $Y_j \notin \mathrm{Span}(X_{s+1} \cup Y_{s+1})$ and so $X_{s+1} \cup Y_{s+1} = \{x_1, \dots, x_{s+1}\} \cup \{y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_t\}$ is a basis for V.

Thus if there is some subset Y_s of Y such that $X_s \cup Y_s$ is a basis, there is a subset Y_{s+1} of Y such that $X_{s+1} \cup Y_{s+1}$ is a basis.

By induction, there is a subset $Y_m = Y'$ of Y such that $X_m \cup Y_m = X \cup Y'$ is a basis. □

Corollary 2.8 *If $X = \{x_1, \dots, x_m\}$ is a frame in a vector space V, and $Y = \{y_1, \dots, y_n\}$ is a span for V, then $m \leq n$.*

Lemma 2.9 *If V is a finitely generated vector space, then any two bases have the same number of vectors, where this number is called the dimension of V, and written* $\dim(V)$.

Proof Let X, Y be two bases with m, n number of elements. Consider X as a frame and Y as a span. Thus $m \leq n$. However Y is also a frame and X a span. Thus $n \leq m$. Hence $m = n$. □

If V' is a vector subspace of a finitely generated vector space V, then any basis for V' can be extended to give a basis for V. To see this, there must exist some finite set $V'' = \{v_1, \dots, v_r\}$ of vectors all belonging to V' such that $\mathrm{Span}(V'') = V'$. Otherwise V could not be finitely generated. As before eliminate members of V'' until a frame is obtained. This gives a basis for V'. Clearly $\dim(V') \leq \dim(V)$. Moreover if V' has a basis $V''' = \{v_1, \dots, v_r\}$ then further linear independent vectors belonging to $V \backslash V'$ can be added to V''' to give a basis for V.

As we showed in Lemma 2.3, the vector space $\Re^n$ has a basis $\{e_1, \dots, e_n\}$ consisting of n elements. Thus $\dim(\Re^n) = n$.

If V^m is a vector subspace of $\Re^n$ of dimension m, where of course $m \leq n$, then in a certain sense V^m is identical to a copy of $\Re^m$ through the origin $\underline{0}$. We make this more explicit below.

2.2 Linear Transformations

In Chap. 1 we considered a morphism from the abelian group $(\Re^2, +)$ to itself. A morphism between vector spaces is called a *linear transformation*.

Definition 2.7 Let V, U be two vector spaces of dimension n, m respectively, over the same field $\mathcal{F}$. Then a *linear transformation* $T : V \to U$ is a function from V to U with domain V, such that

1. for any $a \in \mathcal{F}$, any $v \in V$, $T(\alpha v) = \alpha(T(v))$
2. for any $v_1, v_2 \in V$, $T(v_1 + v_2) = T(v_1) + T(v_2)$.

Note that a linear transformation is simply a morphism between $(V, +)$ and $(U, +)$ which respects the operation of the field $\mathcal{F}$. We shall show that any linear transformation T can be represented by an array of the form

$$M(T) = \begin{pmatrix} a_{11} & & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & & a_{mn} \end{pmatrix}$$

consisting of $n \times m$ elements in F. An array such as this is called an n by m (or $n \times m$) *matrix*. The set of $n \times m$ matrices we shall write as $M(n, m)$.

2.2.1 Matrices

For convenience we shall consider finitely generated vector spaces over $\Re$, so that we restrict attention to linear transformations between $\Re^n$ and $\Re^m$, for any integers n and m. Now let $V = \{v_1, \ldots, v_n\}$ be a basis for $\Re^n$ and $U = \{u_1, \ldots, u_m\}$ a basis for $\Re^m$.

Since V is a basis for $\Re^n$, any vector $x \in \Re^n$ can be written as $x = \sum_{j=1}^n x_j v_j$, with coordinates $(x_1, \ldots, x_n)$.

If T is a linear transformation, then $T(\alpha v_1 + \beta v_2) = T(\alpha v_1) + T(\beta v_2) = \alpha T(v_1) + \beta T(v_2)$. Therefore

$$T(x) = T\left(\sum_{j=1}^n x_j v_j\right) = \sum_{j=1}^n x_j T(v_j).$$

Since each $T(v_j)$ lies in $\Re^m$ we can write $T(v_j) = \sum_{i=1}^m a_{ij} u_i$, where $(a_{1j}, a_{2j}, \ldots, a_{mj})$ are the coordinates of $T(v_j)$ with respect to the basis U for $\Re^m$.

Thus

$$T(x) = \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij} u_i = \sum_{i=1}^m y_i u_i$$

where the ith coordinate, y_i, of $T(x)$ is equal to $\sum_{j=1}^n a_{ij} x_j$.

We obtain a set of linear equations:

$$\begin{array}{lllll} y_1 & = a_{11}x_1 & + a_{12}x_2 & + \cdots a_{1j}x_j & + \cdots a_{1n}x_n \\ \vdots & & \vdots & & \\ y_i & = a_{i1}x_1 & + a_{i2}x_2 & + \cdots a_{ij}x_j & + \cdots a_{in}x_n \\ \vdots & & \vdots & & \\ y_m & = a_{m1}x_1 & + a_{m2}x_2 & + \cdots a_{mj}x_j & + \cdots a_{mn}x_n. \end{array}$$

This set of equations is more conveniently written

$$\begin{array}{cc} \text{row } i & \begin{pmatrix} a_{11} & \dots & a_{ij} & \dots & a_{1n} \\ \vdots & & & & \\ a_{i1} & & a_{ij} & & a_{in} \\ \vdots & & & & \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \\ & j\text{th column} \end{array} \begin{pmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{pmatrix}.$$

or as $M(T)x = y$, where $M(T)$ is the $n \times m$ array whose ith row is $(a_{i1}, \dots, a_{in})$ and whose jth column is $(a_{1j}, \dots, a_{mj})$. This matrix is commonly written as (a_{ij}) where it is understood that $i = 1, \dots, m$ and $j = 1, \dots, n$.

Note that the operation of $M(T)$ on x is as follows: to obtain the ith coordinate, y_i, take the ith row vector $(a_{11}, \dots, a_{1n})$ and form the *scalar product* of this with the column vector $(x_1, \dots, x_n)$, where this scalar product is defined to be $\sum_{j=1}^{n} a_{ij} x_j$.

The coefficients of $T(v_j)$ with respect to the basis $(u_1, \dots, u_m)$ are $(a_{1j}, \dots, a_{mj})$ and these turn up as the jth column of the matrix. Thus we could write the matrix as

$$M(T) = \big(T(v_1) \dots T(v_j) \dots T(v_n)\big)$$

where $T(v_j)$ is the column of coordinates in $\Re^m$. Suppose now that $W = \{w_1, \dots, w_p\}$ is a basis for $\Re^p$ and $S : \Re^m \to \Re^p$ is a linear transformation. Then to represent S as a matrix with respect to the two sets of bases, U and W, for each $i = 1, \dots, m$, we need to know

$$S(u_i) = \sum_{k=1}^{p} b_{ki} w_k.$$

Then as before S is represented by the matrix

$$M(S) = \begin{pmatrix} b_{11} & \dots & b_{1i} & \dots & b_{1m} \\ \vdots & & b_{ki} & & \vdots \\ b_{p1} & \dots & b_{pi} & \dots & b_{pm} \end{pmatrix}$$

where the ith column is the column of coordinates of $S(u_i)$ in $\Re^p$.

We can compute the composition

$$(S \circ T) : \Re^n \xrightarrow{T} \Re^m \xrightarrow{S} \Re^p.$$

The question is how should we compose the two matrices $M(S)$ and $M(T)$ so that the result "corresponds" to the matrix $M(S \circ T)$ which represents $S \circ T$.

First of all we show that $S \circ T : \Re^n \to \Re^p$ is a linear transformation, so that we know that it can be represented by an $(n \times p)$ matrix.

Lemma 2.10 *If $T : \Re^n \to \Re^m$ and $S : \Re^m \to \Re^p$ are linear transformations, then $S \circ T : \Re^n \to \Re^p$ is a linear transformation.*

Proof Consider $\alpha, \beta \in \Re$, $v_1, v_2 \in \Re^n$. Then

$$\begin{aligned}(S \circ T)(\alpha v_1 + \beta v_2) &= S\big[T(\alpha v_1 + \beta v_2)\big] \\ &= S\big(\alpha T(v_1) + \beta T(v_2)\big) \quad \text{since } T \text{ is linear} \\ &= \alpha S\big(T(v_1)\big) + \beta S\big(T(v_2)\big) \quad \text{since } S \text{ is linear} \\ &= \alpha(S \circ T)(v_1) + \beta(S \circ T)(v_2).\end{aligned}$$

Thus $S \circ T$ is linear.

By the previous analysis, $(S \circ T)$ can be represented by an $(n \times p)$ matrix whose jth column is $(S \circ T)(v_j)$. Thus

$$\begin{aligned}(S \circ T)(v_j) &= S\left(\sum_{i=1}^{m} a_{ij} u_i\right) \\ &= \sum_{i=1}^{m} a_{ij} S(u_i) \\ &= \sum_{i=1}^{m} a_{ij} \sum_{k=1}^{p} b_{ki} w_k \\ &= \sum_{k=1}^{p} \left(\sum_{i=1}^{m} a_{ij} b_{ki}\right) w_k.\end{aligned}$$

Thus the kth entry in the jth column of $M(S \circ T)$ is $\sum_{i=1}^{m} b_{ki} a_{ij}$.

Thus $(S \circ T)$ can be represented by the matrix

$$M(S \circ T) = k\text{th row} \begin{pmatrix} & \longleftarrow n \longrightarrow & \\ \cdots & \sum_{i=1}^{m} b_{ki} a_{ij} & \cdots \\ & \\ & j\text{th column} & \end{pmatrix} p$$

The jth column in this matrix can be obtained more simply by operating the matrix $M(S)$ on the jth column vector $T(v_j)$ in the matrix $M(T)$.

Thus $M(S \circ T) = (M(S)(T(v_1)) \dots M(S)(T(v_n))) = M(S) \circ M(T)$.

$$
\begin{aligned}
&\begin{array}{l} k\text{th row of} \\ p \text{ rows}\end{array}
\begin{pmatrix} b_{k1}\ldots & b_{ki}\ldots & b_{km} \\ \longleftarrow & m \text{ columns} & \longrightarrow \end{pmatrix}
\begin{pmatrix} \longleftarrow n \longrightarrow \\ a_{ij} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \\ j\text{th column}\end{pmatrix} m \text{ rows} \\
&= M(S)\circ M(T).
\end{aligned}
$$

□

Thus the "natural" method of matrix composition corresponds to the composition of linear transformations.

Now let $L(\Re^n, \Re^n)$ stand for the set of linear transformations from $\Re^n$ to $\Re^n$. As we have shown, if S, T belong to this set then $S \circ T$ is also a linear transformation from $\Re^n$ to $\Re^n$. Thus composition of functions ($\circ$) is a binary operation $L(\Re^n, \Re^n) \times L(\Re^n, \Re^n) \to L(\Re^n, \Re^n)$.

Let $M : L(\Re^n, \Re^n) \to M(n, n)$ be the mapping which assigns to any linear transformation $T : \Re^n \to \Re^n$ the matrix $M(T)$ as above. Note that M is dependent on the choice of bases $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_n\}$ for the domain and codomain, $\Re^n$. There is in general no reason why these two bases should be the same.

Now let $\circ$ be the method of matrix composition which we have just defined. Thus the mapping M satisfies

$$M(S \circ T) = M(S) \circ M(T)$$

for any two linear transformations, S and T. Suppose now that we are given a linear transformation, $T \in L(\Re^n, \Re^n)$. Clearly the matrix $M(T)$ which represents T with respect to the two bases is unique, and so M is a function.

On the other hand suppose that T, S are both represented by the same matrix $A = (a_{ij})$.

By definition $T(v_j) = S(v_j) = \sum_{i=1}^m a_{ij} u_i$ for each $j = 1, \ldots, n$.

But then $T(x) = S(x)$ for any $x \in \Re^n$, and so $T = S$. Thus M is injective.

Moreover if A is any matrix, then it represents a linear transformation, and so M is surjective. Thus we have a bijective morphism

$$M : \big(L\big(\Re^n, \Re^n\big), \circ\big) \to \big(M(n, n), \circ\big).$$

As we saw in the case of 2×2 matrices, the subset of non-singular matrices in $M(n, n)$ forms a group. We repeat the procedure for the more general case.

2.2.2 The Dimension Theorem

Let $T : V \to U$ be a linear transformation between the vector spaces V, U of dimension n, m respectively over a field $\mathcal{F}$. The transformation is characterised by two subspaces, of V and U.

Definition 2.8

1. the *kernel* of a transformation $T : V \to U$ is the set $\mathrm{Ker}(T) = \{x \in V : T(x) = \underline{0}\}$ in V.
2. The *image* of the transformation is the set $\mathrm{Im}(T) = \{y \in U : \exists\, x \in V \text{ s.t. } T(x) = y\}$.

Both these sets are vector subspaces of U, V respectively. To see this suppose $v_1, v_2 \in \mathrm{Ker}(T)$, and $\alpha, \beta \in F$. Then $T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2) = \underline{0} + \underline{0} = \underline{0}$. Hence $\alpha v_1 + \beta v_2 \in \mathrm{Ker}(T)$.

If $u_1, u_2 \in \mathrm{Im}(T)$ then there exists $v_1, v_2 \in V$ such that $T(v_1) = u_1, T(v_2) = u_2$. But then

$$\begin{aligned}\alpha\mu_1 + \beta\mu_2 &= \alpha T(v_1) + \beta T(v_2)\\ &= T(\alpha v_1 + \beta v_2).\end{aligned}$$

Since V is a vector space, $\alpha v_1 + \beta v_2 \in V$ and so $\alpha u_1 + \beta u_2 \in \mathrm{Im}(T)$.

By the exchange theorem there exists a basis $k_1, \dots, k_p$ for $\mathrm{Ker}(T)$, where $p = \dim \mathrm{Ker}(T)$ and a basis $u_1, \dots, u_s$ for $\mathrm{Im}(T)$ where $s = \dim(\mathrm{Im}(T))$. Here p is called the *kernel rank* of T, often written $kr(T)$, and s is the *rank* of T, or $rk(T)$.

The Dimension Theorem *If $T : V \to U$ is a linear transformation between vector spaces over a field $\mathcal{F}$, where dimension $(V) \neq n$, then the dimension of the kernel and image of T satisfy the relation*

$$\dim\bigl(\mathrm{Im}(T)\bigr) + \dim\bigl(\mathrm{Ker}(T)\bigr) = n.$$

Proof Let $\{u_1, \dots, u_s\}$ be a basis for $\mathrm{Im}(T)$ and for each $i = 1, \dots, s$, let v_i be the vector in V^n such that $T(v_i) = u_i$.

Let v be any vector in V. Then

$$T(v) = \sum_{i=1}^{s} \alpha_i u_i, \quad \text{for } T(v) \in \mathrm{Im}(T).$$

So

$$\begin{aligned}T(v) &= \sum_{i=1}^{s} \alpha_i T(v_i)\\ &= T\left(\sum_{i=1}^{s} \alpha_i v_i\right), \quad \text{and}\end{aligned}$$

$$T\left(v - \sum_{i=1}^{s} \alpha_i v_i\right) = \underline{0},$$

the zero vector in U, *i.e.*, $v - \sum_{i=1}^{s} \alpha_i v_i \in$ kernel T. Let $\{k_1, \dots, k_p\}$ be the basis for $\mathrm{Ker}(T)$.

Then $v - \sum_{i=1}^{s} \alpha_i v_i = \sum_{j=1}^{p} \beta_j k_j$, or $v = \sum_{i=1}^{s} \alpha_i v_i + \sum_{j=1}^{p} \beta_j k_j$. Thus $(v_1, \ldots, v_s, k_1, \ldots, k_p)$ is a span for V.

Suppose we consider

$$\sum_{i=1}^{s} \alpha_i v_i + \sum_{j=1}^{p} \beta_j k_j = \underline{0}. \tag{*}$$

Then, since $T(k_j) = 0$ for $j = 1, \ldots, p$,

$$\begin{aligned} T\left(\sum_{i=1}^{s} \alpha_i v_i + \sum_{j=1}^{p} \beta_j k_j\right) &= \sum_{i=1}^{s} \alpha_i T(v_i) + \sum_{j=1}^{p} \beta_j T(k_j) \\ &= \sum_{i=1}^{s} \alpha_i T(v_i) = \sum_{i=1}^{s} \alpha_i u_i = \underline{0}. \end{aligned}$$

Now $\{u_i, \ldots, u_s\}$ is a basis for $\mathrm{Im}(T)$, and hence these vectors are linearly independent. So $\alpha_i = 0, i = 1, \ldots, s$. Therefore (*) gives $\sum_{j=1}^{p} \beta_j k_j = \underline{0}$.

However $\{k_1, \ldots, k_p\}$ is a basis for $\mathrm{Ker}(T)$ and therefore a frame, so $\beta_j = 0$ for $j = 1, \ldots, p$. Hence $\{v_1, \ldots, v_s, k_1, \ldots, k_p\}$ is a frame, and therefore a basis for V. By the exchange theorem the dimension of V is the unique number of vectors in a basis. Therefore $s + p = n$. □

Note that this theorem is true for general vector spaces. We specialise now to vector spaces $\Re^n$ and $\Re^m$.

Suppose $\{v_1, \ldots, v_n\}$ is a basis for $\Re^n$. The coordinates of v_j with respect to this basis are $(0, \ldots, 1, \ldots, 0)$ with 1 in the jth place. As we have noted the image of v_j under the transformation T can be represented by the jth column $(a_{ij}, \ldots, a_{mj})$ in the matrix $M(T)$, with respect to the original basis $(e_1, \ldots, e_m)$, say, for $\Re^m$. Call the n different column vectors of this matrix $a_1, \ldots, a_j, \ldots, a_n$.

Then the equation $M(T)(x) = y$ is identical to the equation $\sum_{j=1}^{n} x_j a_j = y$ where $x = (x_1, \ldots, x_n)$.

Clearly any vector y in the image of $M(T)$ can be written as a linear combination of the columns $A = \{a_1, \ldots, a_n\}$. Thus $\mathrm{Span}(A) = \mathrm{Im}(M(T))$. Suppose now that A is not a frame. In this case a_n, say, can be written as a linear combination of $\{a_1, \ldots, a_{n-1}\}$, *i.e.*, $\sum_{j=1}^{n} k_{1j} a_j = \underline{0}$ and $k_{1n} \neq 0$. Then the vector $k_1 = (k_{11}, \ldots, k_{1n})$ satisfies $M(T)(k_1) = \underline{0}$. Thus k_1 belongs to $\mathrm{Ker}(M(T))$.

Eliminate a_n, say, and proceed in this way. After p iterations we will have obtained p kernel vectors $\{k_1, \ldots, k_p\}$ and the remaining column vectors $\{a_1, \ldots, a_{n-p}\}$ will form a frame, and thus a basis for the image of $M(T)$.

Consequently $\dim(\mathrm{Im}(M(T))) = n - p = n - \dim(\mathrm{Ker}(M(T)))$. The number of linearly independent columns in the matrix $M(T)$ is called the *rank* of $M(T)$, and is clearly the dimension of the image of $M(T)$. In particular if $M_1(T)$ and $M_2(T)$ are two matrix representations with respect to different bases, of the linear transformation T, then $\mathrm{rank}\, M_1(T) = \mathrm{rank}\, M_2(T) = \mathrm{rank}(T)$.

Thus $\text{rank}(T)$ is an invariant, in the sense of being independent of the particular bases chosen for $\Re^n$ and $\Re^m$.

In the same way the kernel rank of T is an invariant; that is, for any matrix representation $M(T)$ of T we have $\ker\text{rank}(M(T)) = \ker\text{rank}(T)$.

In general if $y \in \text{Im}(T)$, x_0 satisfies $T(x_0) = y$, and k belongs to the kernel, then

$$T(x_0 + k) = T(x_0) + T(k) = y + \underline{0} = y.$$

Thus if x_0 is a solution to the equation $T(x_0) = y$, the point $x_0 + k$ is also a solution. More generally $x_0 + \text{Ker}(T) = \{x_0 + k : k \in \text{Ker}(T)\}$ will also be the set of solutions. Thus for a particular $y \in \text{Im}(T)$, $T^{-1}(y) = \{x : T(x) = y\} = x_0 + \text{Ker}(T)$.

By the dimension theorem $\dim\text{Ker}(T) = n - \text{rank}(T)$. Thus $T^{-1}(y)$ is a geometric object of "dimension" dim $\text{Ker}(T) = n - \text{rank}(T)$.

We defined T to be *injective* iff $T(x_0) = T(x)$ implies $x_0 = x$. Thus T is injective iff $\text{Ker}(T) = \{\underline{0}\}$. In this case, if there is a solution to the equation $T(x_0) = y$, then this solution is unique.

Suppose that $n \leq m$, and that the n different column vectors of the matrix are linearly independent. In this case $\text{rank}(T) = n$ and so $\dim\text{Ker}(T) = 0$. Thus T must be *injective*. In particular if $n < m$ then not every $y \in \Re^m$ belongs to the image of T, and so not every equation $T(x) = y$ has a solution. Suppose on the other hand that $n > m$. In this case the maximum possible rank is m (since n vectors cannot be linearly independent in $\Re^m$ when $n > m$). If $\text{rank}(T) = m$, then there must exist a kernel of dimension $(n - m)$.

Moreover $\text{Im}(T) = \Re^m$, and so for every $y \in \Re^m$ there exists a solution to this equation $T(x) = y$. Thus T is *surjective*. However the solution is not unique, since $T^{-1}(y) = x + \text{Ker}(T)$ is of dimension $(n - m)$ as before.

Suppose now that $n = m$, and that $T : \Re^n \to \Re^n$ has maximal rank n. Then T is both injective and surjective and thus an *isomorphism*. Indeed T will have an inverse *function* $T^{-1} : \Re^n \to \Re^n$. Moreover T^{-1} is linear. To see this note that if $x_1 = T^{-1}(y_1)$ and $x_2 = T^{-1}(y_2)$ then $T(x_1) = y_1$ and $T(x_2) = y_2$ so $T(x_1 + x_2) = y_1 + y_2$. Thus $T^{-1}(y_1 + y_2) = x_1 + x_2 = T^{-1}(y_1) + T^{-1}(y_2)$. Moreover if $x = T^{-1}(\alpha y)$ then $T(x) = \alpha y$. If $\alpha \neq 0$, then $\frac{1}{\alpha}T(x) = T(\frac{1}{\alpha}x) = y$ or $\frac{1}{\alpha}x = T^{-1}(y)$. Hence $x = \alpha T^{-1}(y)$. Thus $T^{-1}(\alpha y) = \alpha T^{-1}(y)$. Since T^{-1} is linear it can be represented by a matrix $M(T^{-1})$. As we know $M : (L(\Re^n, \Re^n), \circ) \to (M(n,n), \circ)$ is a bijective morphism, so M maps the identity linear transformation, Id, to the identity *matrix*

$$M(\text{Id}) = I = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 1 \end{pmatrix}.$$

When T is an isomorphism with inverse T^{-1}, then the representation $M(T^{-1})$ of T^{-1} is $[M(T)]^{-1}$. We now show how to compute the *inverse matrix* $[M(T)]^{-1}$ of an isomorphism.

2.2.3 The General Linear Group

To compute the inverse of an $n \times n$ matrix A, we define, by induction, the determinant of A. For a 1×1 matrix (a_{11}) define det $(A_{11}) = a_{11}$, and for a 2×2 matrix $A = \binom{a_{11}\ a_{12}}{a_{21}\ a_{22}}$ define det $A = a_{11}a_{22} - a_{21}a_{12}$.

For an $n \times n$ matrix A define the (i, j)th *cofactor* to be the determinant of the $(n-1) \times (n-1)$ matrix $A(i, j)$ obtained from A by removing the ith row and jth column, then multiplying by $(-1)^{i+j}$. Write this cofactor as A_{ij}. For example in the 3×3 matrix, the cofactor in the $(1, 1)$ position is

$$A_{11} = det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} = a_{22}a_{33} - a_{32}a_{23}.$$

The $n \times n$ matrix (A_{ij}) is called the *cofactor matrix*.

The determinant of the $n \times n$ matrix A is then $\sum_{j=1}^{n} a_{1j}A_{1j}$. The determinant is also often written as $|A|$.

This procedure allows us to define the determinant of an $n \times n$ matrix. For example if $A = (a_{ij})$ is a 3×3 matrix, then

$$\begin{aligned} |A| &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}). \end{aligned}$$

An alternative way of defining the determinant is as follows. A permutation of n is a bijection $s : \{1, \dots, n\} \to \{1, \dots, n\}$, with degree $d(s)$ the number of exchanges needed to give the permutation.

Then $|A| = \sum_s (-1)^{d(s)} \Pi_{i=1}^{n} a_{is(i)} = a_{11}a_{22}a_{33} \cdots + \cdots$ where the summation is over all permutations. The two definitions are equivalent, and it can be shown that

$$\begin{aligned} |A| &= \sum_{j=1}^{n} a_{ij}A_{ij} \quad \text{(for any } i = 1, \dots, n) \\ &= \sum_{i=1}^{n} a_{ij}A_{ij} \quad \text{(for any } j = 1, \dots, n) \end{aligned}$$

while

$$\begin{aligned} 0 &= \sum_{i=1}^{n} a_{ij}A_{ik} \quad \text{if } j \neq k \\ &= \sum_{j=1}^{n} a_{ij}A_{kj} \quad \text{if } i \neq k. \end{aligned}$$

Thus

$$(a_{ij})\ (A_{jk})^t = \left(\sum_{j=1}^{n} a_{ij} A_{kj}\right)$$
$$= \begin{pmatrix} |A| & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & |A| \end{pmatrix} = |A| I.$$

Here $(A_{jk})^t$ is the $n \times n$ matrix obtained by transposing the rows and columns of (A_{jk}). Now the matrix A^{-1} satisfies $A \circ A^{-1} = I$, and if A^{-1} exists then it is unique. Thus $A^{-1} = \frac{1}{|A|}(A_{ij})^t$.

Suppose that the matrix A is non-singular, so $|A| \neq 0$. Then we can construct an inverse matrix A^{-1}.

Moreover if $A(x) = y$ then $y = A^{-1}(x)$ which implies that A is both injective and surjective. Thus $\operatorname{rank}(A) = n$ and the column vectors of A must be linearly independent.

As we have noted, however, if A is not injective, with $\operatorname{Ker}(A) \neq \{0\}$, then $\operatorname{rank}(A) < n$, and the column vectors of A must be linearly dependent. In this case the inverse A^{-1} is not a function and cannot therefore be represented by a matrix and so we would expect $|A|$ to be zero.

Lemma 2.11 *If A is an $n \times n$ matrix with* $\operatorname{rank}(A) < n$ *then* $|A| = 0$.

Proof Let A' be the matrix obtained from A by adding a multiple (α) of the kth column of A to the jth column of A. The jth column of A' is therefore $a_j + \alpha a_k$. This operation leaves the jth column of the cofactor matrix unchanged. Thus

$$\begin{aligned} |A'| &= \sum_{i=1}^{n} a'_{ij} A_{ij} \\ &= \sum_{i=1}^{n} (a_{ij} + \alpha a_{ik}) A_{ij} \\ &= \sum_{i=1}^{n} a_{ij} A_{ij} + \alpha \sum_{i=1}^{n} a_{ik} A_{ij} \\ &= |A| + 0 = |A|. \end{aligned}$$

Suppose now that the columns of A are linearly dependent, and that $a_j = \sum_{k \neq j} \alpha_k a_k$ for example. Let A' be the matrix obtained from A by substituting $a'_j = 0 = a_j - \sum_{k \neq j} \alpha_k a_k$ for the jth column.

By the above $|A'| = \sum_{i=1}^{n} a'_{ij} A_{ij} = 0 = |A|$. □

Suppose now that A, B are two non-singular matrices (a_{ij}), (b_{ki}). The composition is then $B \circ A = (\sum_{i=1}^{m} b_{ki} a_{ij})$ with determinant

$$|B \circ A| = \sum_{s} (-1)^{d(s)} \prod_{k=1}^{n} \circ \left(\sum_{i=1}^{m} b_{ki} a_{is(k)} \right).$$

This expression can be shown to be equal to

$$\sum_{s} (-1)^{d(s)} \prod_{i=1}^{n} a_{is(i)} \sum_{s} (-1)^{d(s)} \prod_{i=1}^{n} b_{ks(k)} = |B||A| \neq 0.$$

Hence the composition $(B \circ A)$ has an inverse $(B \circ A)^{-1}$ given by $A^{-1} \circ B^{-1}$.

Now let $(GL(\Re^n, \Re^n), \circ)$ be the set of invertible linear transformations, with $\circ$ composition of functions, and let $M^*(n, n)$ be the set of non-singular $n \times n$ matrices. Choice of bases $\{v_1, \dots, v_n\}, \{u_1, \dots, u_n\}$ for the domain and codomain defines a morphism

$$M : \left(GL\left(\Re^n, \Re^n\right), \circ\right) \to \left(M^*(n, n), \circ\right).$$

Suppose now that T belongs to $GL(\Re^n, \Re^n)$. As we have seen this is equivalent to $|M(T)| \neq 0$, so the image of M is precisely $M^*(n, n)$. Moreover if $|M(T)| \neq 0$ then $|M(T^{-1})| = \frac{1}{|M(T)|}$ and $M(T^{-1})$ belongs to $M^*(n, n)$. On the other hand if $S, T \in GL(\Re^n, \Re^n)$ then $S \circ T$ also has rank n, and has inverse $T^{-1} \circ S^{-1}$ with rank n.

The matrix $M(S \circ T)$ representing $T \circ S$ has inverse

$$\begin{aligned} M\left(T^{-1} \circ S^{-1}\right) &= M\left(T^{-1}\right) \circ M\left(S^{-1}\right) \\ &= \left[M(T)\right]^{-1} \circ \left[M(S)\right]^{-1}. \end{aligned}$$

Thus M is an *isomorphism* between the two groups $(GL(\Re^n, \Re^n, \circ))$ and $(M^*(n, n), \circ)$.

The group of invertible linear transformations is also called the *general linear group*.

2.2.4 Change of Basis

Let $L(\Re^n, \Re^m)$ stand for the set of linear transformations from $\Re^n$ to $\Re^m$, and let $M(n, m)$ stand for the set of $n \times m$ matrices. We have seen that the choice of bases for $\Re^n, \Re^m$ defines a *function*.

$$M : L\left(\Re^n, \Re^m\right) \to M(n, m)$$

which take a linear transformation T to its representation $M(T)$. We now examine the relationship between two representations $M_1(T)$, $M_2(T)$ of a single linear transformation.

Basis Change Theorem *Let $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_m\}$ be bases for $\Re^n, \Re^m$ respectively.*

Let T be a linear transformation which is represented by a matrix $A = (a_{ij})$ with respect to these bases. If $V' = \{v'_1, \ldots, v'_n\}$, $U' = \{u'_1, \ldots, u'_m\}$ are new bases for $\Re^n, \Re^m$ then T is represented by the matrix $B = Q^{-1} \circ A \circ P$, where P, Q are respectively $(n \times n)$ and $(m \times m)$ invertible matrices.

Proof For each $v'_k \in V' = \{v'_1, \ldots, v'_n\}$ let $v'_k = \sum_{i=1}^n b_{ik} v_i$ and $b_k = (b'_{1k}, \ldots, b_{nk})$.

Let $P = (b_1, \ldots, b_n)$ where the kth column of P is the column of coordinates of b_k. With respect to the new basis V', v'_k has coordinates $e_k = (0, \ldots, 1, \ldots, 0)$ with a 1 in the kth place.

But then $P(e_k) = b_k$ the coordinates of v'_k with respect to V.

Thus P is the matrix that transforms coordinates with respect to V' into coordinates with respect to V. Since V is a basis, the columns of P are linearly independent, and so rank $P = n$, and P is invertible.

In the same way let $u'_k = \sum_{i=1}^m c_{ik} u_i$, $c_k = (c_{1k}, \ldots, c_{mk})$ and $Q = (c_1, \ldots, c_m)$ the matrix with columns of these coordinates.

Hence Q represents change of basis from U' to U. Since Q is an invertible $m \times m$ matrix it has inverse Q^{-1} which represents change of basis from U to U'.

Thus we have the diagram

$$\begin{array}{ccc} \{v_1, \ldots, v_n\} & \xrightarrow{A} & \{u_1, \ldots, u_m\} \\ P \uparrow & & Q^{-1} \downarrow \uparrow Q \\ \{v'_1, \ldots, v'_n\} & \xrightarrow{B} & \{u'_1, \ldots, u'_m\} \end{array}$$

from which we see that the matrix B, representing the linear transformation $T : \Re^n \to \Re^m$ with respect to the new bases is given by $B = Q^{-1} \circ A \circ P$. □

Isomorphism Theorem *Any linear transformation $T : \Re^n \to \Re^m$ of rank r can be represented, by suitable choice of bases for $\Re^n$ and $\Re^m$, by an $n \times m$ matrix*

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \quad \textit{where } I_r = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \textit{ is the } (r \times r) \textit{ identity matrix.}$$

In particular

1. *if $n < m$ and T is injective then there is an isomorphism $S : \Re^m \to \Re^m$ such that $S \circ T(x_1, \ldots, x_n) = (x_1, \ldots, x_n, 0, \ldots, 0)$ with $(n - m)$ zero entries, for any vector $(x_1, \ldots, x_n)$ in $\Re^n$*

2. *if $n \geq m$ and T is surjective then there are isomorphisms $\Re : \Re^n \to \Re^n$, $S : \Re^m \to \Re^m$ such that $S \circ T \circ R(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$. If $n = m$, then $S \circ T \circ R$ is the identity isomorphism.*

Proof Of necessity $\text{rank}(T) = r \leq \min(n, m)$. If $r < n$, let $p = n - r$ and choose a basis $k_1, \ldots, k_p$ for $\text{Ker}(T)$. Let $V = \{v_1, \ldots, v_n\}$ be the original basis for $\Re^n$. By the exchange theorem there exists $r = (n - p)$ different members $\{v_1, \ldots, v_r\}$ say of V such that $V' = \{v_1, \ldots, v_r, k_1, \ldots, k_p\}$ is a basis for $\Re^n$.

Choose V' as the new basis for $\Re^n$, and let P be the basis change matrix whose columns are the column vectors in V'. As in the proof of the dimension theorem the image of the vectors $v_1, \ldots, v_{n-p}$ under T provide a basis for the image of T. Let $U = \{u_1, \ldots, u_m\}$ be the original basis of $\Re^m$. By the exchange theorem there exists some subset $U' = \{u_1, \ldots, u_{m-r}\}$ of U such that $U'' = \{T(v_1), \ldots, T(v_r), u_1, \ldots, u_{m-r}\}$ form a basis for $\Re^m$. Note that $T(v_1), \ldots, T(v_r)$ are represented by the r linearly independent columns of the original matrix A representing T. Now let Q be the matrix whose columns are the members of U''. By the basis change theorem, $B = Q^{-1} \circ A \circ P$, where B is the matrix representing T with respect to these new bases. Thus we obtain

$$\begin{array}{ccc}
\{v_1, \ldots, v_n\} & \xrightarrow{A} & \{u_1, \ldots, u_m\} \\
P \uparrow & & Q^{-1} \downarrow \\
\{v_1, \ldots, v_r, k_1 \ldots\} & \xrightarrow{B} & \left\{T(v_1) \ldots T(v_r), u_1, \ldots, u_{m-r}\right\}.
\end{array}$$

With respect to these new bases, the matrix B representing T has the required form: $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

1. If $n < m$ and T is injective then $r = n$. Hence P is the identity matrix, and so $B = Q^{-1} \circ A$.

   But Q^{-1} is an $m \times m$ invertible matrix, and thus represents an isomorphism $\Re^n \longrightarrow \Re^n$, while

$$B \begin{pmatrix} x_1 \\ x_n \end{pmatrix} = \begin{pmatrix} I_n \\ 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ \vdots \\ 0 \end{pmatrix}.$$

   Write a vector $x = \sum_{i=1}^{n} x_i v_i$ as $(x_1, \ldots, x_n)$, and let S be the linear transformation $\Re^m \to \Re^m$ represented by the matrix Q^{-1}. Then $S \circ T(x_1, \ldots, x_n) = (x_1, \ldots, x_n, 0, \ldots, 0)$.
2. If $n \geq m$ and T is surjective then $\text{rank}(T) = m$, and $\dim \text{Ker}(T) = n - m$. Thus $B = (I_m\ 0) = Q^{-1} \circ A \circ P$. Let S, R be the linear transformations represented by Q^{-1} and P respectively.

Then $S \circ T \circ R(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$. If $n = m$ then $S \circ T \circ R$ is the identity transformation. □

Suppose now that V, U are the two bases for $\Re^n, \Re^m$ as in the basis theorem. A linear transformation $T : \Re^n \longrightarrow \Re^m$ is represented by a matrix $M_1(T)$ with respect to these bases. If V', U' are two new bases, then T will be represented by the matrix $M_2(T)$, and by the basis theorem

$$M_2(T) = Q^{-1} \circ M_1(T) \circ P$$

where Q, P are non-singular $(m \times m)$ and $(n \times n)$ matrices respectively. Since $M_1(T)$ and $M_2(T)$ represent the same linear transformation, they are in some sense equivalent. We show this more formally.

Say the two matrices $A, B \in M(n,m)$ are *similar* iff there exist non singular square matrices $P \in M^*(n,n)$ and $Q \in M^*(m,m)$ such that $B = Q^{-1} \circ A \circ P$, and in this case write $B \sim A$.

Lemma 2.12 *The similarity relation* $(\sim)$ *on* $M(n,m)$ *is an equivalence relation.*

Proof

1. To show that $\sim$ is reflexive note that $A = I_m^{-1} \circ A \circ I_n$ where I_m, I_n are respectively the $(m \times m)$ and $(n \times n)$ identity matrices.
2. To show that $\sim$ is symmetric we need to show that $B \sim A$ implies that $A \sim B$. Suppose therefore that $B = Q^{-1} \circ A \circ P$.
Since $Q \in M^*(m,m)$ it has inverse $Q^{-1} \in M^*(m,m)$.
Moreover $(Q^{-1})^{-1} \circ Q^{-1} = I_m$, and thus $Q = (Q^{-1})^{-1}$. Thus

$$\begin{aligned} Q \circ B \circ P^{-1} &= \left(Q \circ Q^{-1}\right) \circ A \circ \left(P \circ P^{-1}\right) \\ &= A \\ &= \left(Q^{-1}\right)^{-1} \circ B \circ \left(P^{-1}\right). \end{aligned}$$

Thus $A \sim B$.
3. To show $\sim$ is transitive, we seek to show that $C \sim B \sim A$ implies $C \sim A$. Suppose therefore that $C = R^{-1} \circ B \circ S$ and $B = Q^{-1} \circ A \circ P$, where $R, Q \in M^*(m,m)$ and $S, P \in M^*(n,n)$. Then

$$\begin{aligned} C &= \left(R^{-1} \circ Q^{-1}\right) \circ A \circ P \circ S \\ &= (Q \circ R)^{-1} \circ A \circ (P \circ S). \end{aligned}$$

Now $(M^*(m,m), \circ)$, $(M^*(n,n), \circ)$ are both groups and so $Q \circ R \in M^*(m,m)$, $P \circ S \in M^*(n,n)$. Thus $C \sim A$. □

The isomorphism theorem shows that if there is a linear transformation $T : \Re^n \longrightarrow \Re^m$ of rank r, then the $(n \times m)$ matrix $M_1(T)$ which represents T, with respect to some pair of the bases, is similar to an $n \times m$ matrix

$$B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \quad i.e., \ M(T) \sim B.$$

If S is a second linear transformation of rank r then $M_1(S) \sim B$.

By Lemma 2.12, $M_1(S) \sim M_1(T)$.

Suppose now that U', V' are a second pair of bases for $\Re^n, \Re^m$ and let $M_2(S), M_2(T)$ represent S and T. Clearly $M_2(S) \sim M_2(T)$.

Thus if S, T are linear transformations $\Re^m \longrightarrow \Re^n$ we may say that S, T are equivalent iff for any choice of bases the matrices $M(S), M(T)$ which represent S, T are similar.

For any linear transformation $T \in L(\Re^n, \Re^m)$ let $[T]$ be the equivalence class $\{S \in L(\Re^n, \Re^m) : S \sim T\}$. Alternatively a linear transformation S belongs to $[T]$ iff $\text{rank}(S) = \text{rank}(T)$. Consequently the equivalence relation partitions $L(\Re^n, \Re^m)$ into a finite number of distinct equivalence classes where each class is classified by its rank, and the rank runs from 0 to $\min(n, m)$.

2.2.5 Examples

Example 2.1 To illustrate the use of these procedures in the solution of linear equations, consider the case with $n < m$ and the equation $A(x) = y$ where

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 5 & 0 & 3 \\ -1 & -4 & 5 \\ 3 & 2 & -1 \end{pmatrix} \quad \text{and} \quad y_1 = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \qquad y_2 = \begin{pmatrix} 0 \\ 5 \\ -5 \\ 5 \end{pmatrix}.$$

To find $\text{Im}(A)$, we first of all find $\text{Ker}(A)$. The equation $A(x) = \underline{0}$ gives four equations

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 0 \\ 5x_1 + 0 + 3x_3 &= 0 \\ -x_1 - 4x_2 + 5x_3 &= 0 \\ 3x_1 + 2x_2 - x_3 &= 0 \end{aligned}$$

with solution $k = (x_1, x_2, x_3) = (-3, 7, 5)$.

Thus $\text{Ker}(A) \supset \{\lambda k \in \Re^3 : \lambda \in \Re\}$. Hence $\dim \text{Im}(A) \leq 2$. Clearly the first two columns (a_1, a_2) of A are linearly independent and so $\dim \text{Im}(A) = 2$. However $y_2 = a_1 + a_2$. Thus a particular solution to the equation $A(x) = y_2$ is $x_0 = (1, 1, 0)$.

The full set of solutions to the equation is

$$x_0 + \mathrm{Ker}(A) = \big\{(1, 1, 0) + \lambda(-3, 7, 5) : \lambda \in \Re\big\}.$$

To see whether $y_1 \in \mathrm{Im}(A)$ we need only attempt to solve the equation $y_1 = \alpha a_1 + \beta a_2$. This gives

$$\begin{aligned} -1 &= \alpha - \beta \\ 1 &= 5\alpha \\ -1 &= -\alpha - 4\beta \\ 1 &= 3\alpha + 2\beta. \end{aligned}$$

From the first two equations $\alpha = \frac{1}{5}$, $\beta = \frac{6}{5}$, which is incompatible with the fourth equation. Thus y_1 cannot belong to $\mathrm{Im}(A)$.

Example 2.2 Consider now an example of the case $n > m$, where

$$A = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & 1 & 1 \end{pmatrix} : \Re^5 \longrightarrow \Re^2.$$

Obviously the first two columns are linearly independent and so $\dim \mathrm{Im}(A) \geq 2$. Let $\{a_i : i = 1, \ldots, 5\}$ be the five column vectors of the matrix and consider the equation

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus $k_1 = (1, -1, -1, 0, 0)$ belongs to $\mathrm{Ker}(A)$. On the other hand

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus $k_2 = (1, 1, 0, -3, 0)$ and $k_3 = (1, 1, 0, 0, -3)$ both belong to $\mathrm{Ker}(A)$.

Consequently the rank of A has its maximal value of 2, while the kernel is three-dimensional. Hence for any $y \in \Re^2$ there is a set of solutions of the form $x_0 + \mathrm{Span}\{k_1, k_2, k_3\}$ to the equation $A(x) = y$.

Change the bases of $\Re^5$ and $\Re^2$ to

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \\ 0 \\ -3 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -3 \end{pmatrix}$$

and $\binom{2}{1}$, $\binom{1}{2}$ respectively, then

$$B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & 1 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Example 2.3 Consider the matrix

$$Q = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 5 & 0 & 0 & 0 \\ -1 & -4 & 1 & 0 \\ 3 & 2 & 0 & 1 \end{pmatrix}.$$

Since $|Q| = 5$ we can compute its inverse. The cofactor matrix (Q_{ij}) of Q is

$$\begin{pmatrix} 0 & -5 & -20 & 10 \\ 1 & 1 & 5 & -5 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

and thus

$$Q^{-1} = \frac{1}{|Q|}(Q_{ij})^t = \begin{pmatrix} 0 & \frac{1}{5} & 0 & 0 \\ -1 & \frac{1}{5} & 0 & 0 \\ -4 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{pmatrix}.$$

Example 2.4 Let $T : \Re^3 \to \Re^4$ be the linear transformation represented by the matrix A of Example 2.1, with respect to the standard bases for $\Re^3, \Re^4$. We seek to change the bases so as to represent T by a diagonal matrix

$$B = \begin{pmatrix} I_r & \\ & 0 \end{pmatrix}.$$

By Example 2.1, the kernel is spanned by $(-3, 7, 5)$, and so we choose a new basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad k = \begin{pmatrix} -3 \\ 7 \\ 5 \end{pmatrix}$$

with basis change matrix $P=(e_1,e_2,k)$. Note that $|P|=5$ and P is non-singular. Thus $\{e_1,e_2,k\}$ form a basis for $\Re^3$. Now $\mathrm{Im}(A)$ is spanned by the first two columns a_1,a_2, of A. Moreover $A(e_1)=a_1$ and $A(e_2)=a_2$. Thus choose

$$a_1=\begin{pmatrix}1\\5\\-1\\3\end{pmatrix},\qquad a_2=\begin{pmatrix}-1\\0\\-4\\2\end{pmatrix},\qquad e_3'=\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\qquad e_4'=\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

as the new basis for $\Re^4$. Let $Q=(a_1,a_2,e_3',e_4')$ be the basis change matrix. The inverse Q^{-1} is computed in Example 2.3. Thus we have $(B)=Q^{-1}\circ A\circ P$.

To check that this is indeed the case we compute:

$$\begin{aligned}Q^{-1}\circ A\circ P&=\begin{pmatrix}0&\frac15&0&0\\-1&\frac15&0&0\\-4&1&1&0\\2&-1&0&1\end{pmatrix}\begin{pmatrix}1&1&2\\5&0&3\\-1&-4&5\\3&2&-1\end{pmatrix}\begin{pmatrix}1&0&-3\\0&1&7\\0&0&5\end{pmatrix}\\&=\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\\0&0&0\end{pmatrix}\end{aligned}$$

as required.

2.3 Canonical Representation

When considering a linear transformation $T:\Re^n\longrightarrow\Re^n$ it is frequently convenient to change the basis of $\Re^n$ to a new basis $V=\{v_1,\dots,v_n\}$ such that T is now represented by a matrix

$$M_2(T)=P^{-1}\circ M_1(T)\circ P.$$

In this case it is generally not possible to obtain $M_2(T)$ in the form $\left(\begin{smallmatrix}I_r&0\\0&0\end{smallmatrix}\right)$ as before.

Under certain conditions however $M_2(T)$ can be written in a diagonal form

$$\begin{pmatrix}\lambda_1&&0\\&\cdot&\\0&&\lambda_n\end{pmatrix},$$

where $\lambda_1,\dots,\lambda_n$ are known as the *eigenvalues*.

More explicitly, a vector x is called an *eigenvector* of the matrix A iff there is a solution to the equation $A(x)=\lambda x$ where λ is a real number. In this case, λ is called the *eigenvalue* associated with the eigenvector x. (Note that we assume $x\neq\underline{0}$.)

2.3.1 Eigenvectors and Eigenvalues

Suppose that there are n linearly independent eigenvectors $\{x_1, \ldots, x_n\}$ for A, where (for each $i = 1, \ldots, n$) λ_i is the eigenvalue associated with x_i. Clearly the eigenvector x_i belongs to Ker(A) iff $\lambda_i = 0$. If rank$(A) = r$ then there would a subset $\{x_1, \ldots, x_r\}$ of eigenvectors which form a basis for Im(A), while $\{x_1, \ldots, x_n\}$ form a basis for $\Re^n$. Now let Q be the $(n \times n)$ matrix representing a basis change from the new basis to the original basis. That is to say the ith column, v_i, of Q is the coordinate of x_i with respect to the original basis.

After transforming, the original becomes

$$Q^{-1} \circ A \circ Q = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \lambda_r & \vdots \\ 0 & & 0 \end{pmatrix} = \wedge,$$

where rank $\wedge =$ rank $A = r$.

In general we can perform this diagonalisation only if there are enough eigenvectors, as the following lemma indicates.

Lemma 2.13 *If A is an $n \times n$ matrix, then there exists a non-singular matrix Q, and a diagonal matrix $\wedge$ such that $\wedge = Q^{-1}AQ$ iff the eigenvectors of A form a basis for $\Re^n$.*

Proof

1. Suppose the eigenvectors form a basis, and let Q be the eigenvector matrix. By definition, if v_i is the ith column of Q, then $A(v_i) = \lambda_i v_i$, where λ_i is real. Thus $AQ = Q\wedge$. But since $\{v_1, \ldots, v_n\}$ is a basis, Q^{-1} exists and so $\wedge = Q^{-1}AQ$.
2. On the other hand if $\wedge = Q^{-1}AQ$, where Q is non-singular then $AQ = Q\wedge$. But this is equivalent to $A(v_1) = \lambda_i v_i$ for $i = 1, \ldots, n$ where λ_i is the ith diagonal entry in $\wedge$, and v_i is the ith column of Q.
   Since Q is non-singular, the columns $\{v_1, \ldots, v_n\}$ are linearly independent, and thus the eigenvectors form a basis for $\Re^n$. □

If there are n distinct (real) eigenvalues then this gives a basis, and thus a diagonalisation.

Lemma 2.14 *If $\{v_1, \ldots, v_m\}$ are eigenvectors corresponding to distinct eigenvalues $\{\lambda_1, \ldots, \lambda_m\}$, of a linear transformation $T : \Re^n \to \Re^n$, then $\{v_1, \ldots, v_m\}$ are linearly independent.*

Proof Since v_1 is assumed to be an eigenvector, it is non-zero, and thus $\{v_1\}$ is a linearly independent set. Proceed by induction.

Suppose $V_k = \{v_1, \ldots, v_k\}$, with $k < m$, are linearly independent. Let v_{k+1} be another eigenvector and suppose

$$v = \sum_{r=1}^{k+1} a_r v_r = \underline{0}.$$

Then $\underline{0} = T(v) = \sum_{r=1}^{k+1} a_r T(v_r) = \sum_{r=1}^{k+1} a_r \lambda_r v_r$.

If $\lambda_{k+1} = 0$, then $\lambda_i \neq 0$ for $i = 1, \ldots, k$ and by the linear independence of V_k, $a_r \lambda_r = 0$, and thus $a_r = 0$ for $r = 1, \ldots, k$.

Suppose $\lambda_{k+1} \neq 0$. Then

$$\lambda_{k+1} v = \sum_{r=1}^{k+1} \lambda_{k+1} a_r v_r = \sum_{r=1}^{k+1} a_r \lambda_r v_r = \underline{0}.$$

Thus $\sum_{r=1}^{k} (\lambda_{k+1} - \lambda_r) a_r v_r = \underline{0}$.

By the linear independence of V_k, $(\lambda_{k+1} - \lambda_r) a_r = 0$ for $r = 1, \ldots, k$.

But the eigenvalues are distinct and so $a_r = 0$, for $r = 1, \ldots, k$.

Thus $a_{k+1} v_{k+1} = \underline{0}$ and so $a_r = 0, r = 1, \ldots, k+1$. Hence

$$V_{k+1} = \{v_1, \ldots, v_{k+1}\}, \quad k < m,$$

is linearly independent.

By induction V_m is a linearly independent set. □

Having shown how the determination of the eigenvectors gives a diagonalisation, we proceed to compute eigenvalues.

Consider again the equation $A(x) = \lambda x$. This is equivalent to the equation $A'(x) = 0$, where

$$A' = \begin{pmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdot & \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & & & a_{nn} - \lambda \end{pmatrix}.$$

For this equation to have a non zero solution it is necessary and sufficient that $|A'| = 0$. Thus we obtain a polynomial equation (called the characteristic equation) of degree n in λ, with n roots $\lambda_1, \ldots, \lambda_n$ not necessarily all real. In the 2×2 case for example this equation is $\lambda^2 - \lambda(a_{11} + a_{22}) + (a_{11}a_{22} - a_{21}a_{12}) = 0$. If the roots of this equation are λ_1, λ_2 then we obtain

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - \lambda(\lambda_1 + \lambda_2) \; + \lambda_1 \lambda_2.$$

Hence

$$\lambda_1\lambda_2 = (a_{11}a_{22} - a_{21}a_{22}) = |A|$$
$$\lambda_1 + \lambda_2 = a_{11} + a_{22}.$$

The sum of the diagonal elements of a matrix is called the *trace* of A. In the 2×2 case therefore

$$\lambda_1\lambda_2 = |A|, \qquad \lambda_1 + \lambda_2 = a_{11} + a_{22} = \text{trace}(A).$$

In the 3×3 case we find

$$\begin{aligned}
&(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \\
&\quad = \lambda^3 - \lambda^2(\lambda_1 + \lambda_2 + \lambda_3) + \lambda(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3) - \lambda_1\lambda_2\lambda_3 \\
&\quad = \lambda^3 - \lambda^2(\text{trace}\, A) + \lambda(A_{11} + A_{22} + A_{33}) - |A| = 0,
\end{aligned}$$

where A_{ii} is the ith diagonal cofactor of A. Suppose all the roots are non-zero (this is equivalent to the non-singularity of the matrix A). Let

$$\wedge = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

be the diagonal eigenvalue matrix, with $|\wedge| = \lambda_1\lambda_2\lambda_3$.

The cofactor matrix of $\wedge$ is then

$$\begin{pmatrix} \lambda_2\lambda_3 & 0 & 0 \\ 0 & \lambda_1\lambda_3 & 0 \\ 0 & 0 & \lambda_1\lambda_2 \end{pmatrix}.$$

Thus we see that the sum of the diagonal cofactors of A and $\wedge$ are identical. Moreover trace $(A) =$ trace $(\wedge)$ and $|\wedge| = |A|$.

Now let $\sim$ be the equivalence relation defined on $L(\Re^n, \Re^n)$ by $B \sim A$ iff there exist basis change matrices P, Q and a diagonal matrix $\wedge$ such that

$$\wedge = P^{-1}AP = Q^{-1}BQ.$$

On the set of matrices which can be diagonalised, $\sim$ is an equivalence relation, and each class is characterised by n invariants, namely the trace, the determinant, and $(n-2)$ other numbers involving the cofactors.

2.3.2 Examples

Example 2.5 Let

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

The characteristic equation is

$$\begin{aligned}(2-\lambda)\big[(1-\lambda)(-2-\lambda)\big] - 1(-2) - (-2(1-\lambda) &= -\lambda\big(\lambda^2 - \lambda - 2\big) \\ &= -\lambda(\lambda-2)(\lambda+1) \\ &= 0.\end{aligned}$$

Hence $(\lambda_1, \lambda_2, \lambda_3) = (0, 2, -1)$. Note that $\lambda_1 + \lambda_2 + \lambda_3 = \text{trace}(A) = 1$ and

$$\lambda_2\lambda_3 = -2 = A_{11} + A_{22} + A_{33}.$$

Eigenvectors corresponding to these eigenvalues are

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \qquad x_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \qquad x_3 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Let P be the basis change matrix given by these three column vectors. The inverse can be readily computed, to give

$$P^{-1}AP = \begin{pmatrix} 1 & -1 & -1 \\ \frac{1}{3} & \frac{1}{3} & 0 \\ -\frac{2}{3} & \frac{1}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -1 \\ 0 & 1 & 1 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Suppose we now compute $A^2 = A \circ A : \Re^3 \to \Re^3$. This can easily be seen to be

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

The characteristic function of A^2 is $(\lambda^3 - 5\lambda^2 + 4\lambda)$ with roots $\mu_1 = 0$, $\mu_2 = 4$, $\mu_3 = 1$.

In fact the eigenvectors of A^2 are x_1, x_2, x_3, the same as A, but with eigenvalues $\lambda_1^2, \lambda_2^2, \lambda_3^2$. In this case $\text{Im}(a) = \text{Im}(A^2)$ is spanned by $\{x_2, x_3\}$ and $\text{Ker}(A) = \text{Ker}(A^2)$ has basis $\{x_1\}$.

More generally consider a linear transformation $A : \Re^n \to \Re^n$. Then if x is an eigenvector with a non-zero eigenvalue λ, $A^2(x) = A \circ A(x) = A[\lambda x] = \lambda A(x) = \lambda^2 x$, and so $x \in \text{Im}(A) \cap \text{Im}(A^2)$.

If there exist n *distinct* real roots to the characteristic equation of A, then a basis consisting of eigenvectors can be found. Then A can be diagonalized, and $\mathrm{Im}(A) = \mathrm{Im}(A^2)$, $\mathrm{Ker}(A) = \mathrm{Ker}(A^2)$.

Example 2.6 Let

$$A = \begin{pmatrix} 3 & -1 & -1 \\ 1 & 3 & -7 \\ 5 & -3 & 1 \end{pmatrix}$$

Then $\mathrm{Ker}(A)$ has basis $\{(1, 2, 1)\}$, and $\mathrm{Im}(A)$ has basis $\{(3, 1, 5), (-1, 3, -3)\}$. The eigenvalues of A are $0, 0, 7$. Since we cannot find three linearly independent eigenvectors, A cannot be diagonalised. Now

$$A^2 = \begin{pmatrix} 3 & -3 & 3 \\ -29 & 29 & -29 \\ 17 & -17 & 17 \end{pmatrix}$$

and thus $\mathrm{Im}(A^2)$ has basis $\{(3, -29, 17)\}$. Note that

$$\begin{pmatrix} 3 \\ -29 \\ 17 \end{pmatrix} = -2 \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} - 9 \begin{pmatrix} -1 \\ 3 \\ -3 \end{pmatrix} \in \mathrm{Im}(A)$$

and so $\mathrm{Im}(A^2)$ is a *subspace* of $\mathrm{Im}(A)$.

Moreover $\mathrm{Ker}(A^2)$ has basis $\{(1, 2, 1), (1, -1, 0)\}$ and so $\mathrm{Ker}(A)$ is a subspace of $\mathrm{Ker}(A^2)$.

This can be seen more generally. Suppose $f : \Re^n \to \Re^n$ is linear, and $x \in \mathrm{Ker}(f)$. Then $f^2(x) = f(f(x)) = \underline{0}$, and so $x \in \mathrm{Ker}(f^2)$. Thus $\mathrm{Ker}(f) \subset \mathrm{Ker}(f^2)$. On the other hand if $v \in \mathrm{Im}(f^2)$ then there exists $w \in \Re^n$ such that $f^2(w) = v$. But $f(w) \in \Re^n$ and so $f(f(w)) = v \in \mathrm{Im}(f)$. Thus $\mathrm{Im}(f^2) \subset \mathrm{Im}(f)$.

2.3.3 Symmetric Matrices and Quadratic Forms

Given two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\Re^n$, let $\langle x, y \rangle = \sum_{i=1}^n x_i y_i \in \Re$ be the scalar product of x and y. Note that $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle = \langle x, \lambda y \rangle$ for any real λ. (We use $\langle -, - \rangle$ to distinguish the scalar product from a vector in $\Re^2$. However the notations (x, y) or $x \cdot y$ are often used for scalar product.)

An $n \times n$ matrix $A = (a_{ij})$ may be regarded as a map $A^* : \Re^n \times \Re^n \to \Re$, where $A^*(x, y) = \langle x, A(y) \rangle$.

A^* is linear in both x and y and is called *bilinear*. By definition $\langle x, A(y) \rangle = \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} y_j$.

Call an $n \times n$ matrix A *symmetric* iff $A = A^t$ where $A^t = (a_{ji})$ is obtained from A by exchanging rows and columns.

In this case $\langle A(x), y\rangle = \sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ji}\ x_i)y_j = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\ a_{ij}y_j$, since $a_{ij} = a_{ji}$ for all i, j.

Hence $\langle A(x), y\rangle = \langle x, A(y)\rangle$ for any $x, y \in \Re^n$ whenever A is symmetric.

Lemma 2.15 *If A is a symmetric $n \times n$ matrix, and x, y are eigenvectors of A corresponding to distinct eigenvalues then $\langle x, y\rangle = 0$, i.e., x and y are orthogonal.*

Proof Let $\lambda_1 \neq \lambda_2$ be the eigenvalues corresponding to the distinct eigenvectors x, y. Now

$$\begin{aligned}\langle A(x), y\rangle &= \langle x, A(y)\rangle\\ &= \langle \lambda_1 x, y\rangle = \langle x, \lambda_2 y\rangle\\ &= \lambda_1\langle x, y\rangle = \lambda_2\langle x, y\rangle.\end{aligned}$$

Here $\langle A(x), y\rangle = \langle x, A(y)\rangle$ since A is symmetric. Moreover $\langle x, \lambda y\rangle = \sum_{i=1}^{n} x_i(\lambda y_i) = \lambda\langle x, y\rangle$. Thus $(\lambda_1 - \lambda_2)\langle x, y\rangle = 0$. If $\lambda_1 \neq \lambda_2$ then $\langle x, y\rangle = 0$. □

Lemma 2.16 *If there exist n distinct eigenvalues to a symmetric $n \times n$ matrix A, then the eigenvectors $X = \{x_1, \ldots, x_n\}$ form an orthogonal basis for $\Re^n$.*

Proof Directly by Lemmas 2.14 and 2.15.

We may also give a brief direct proof of Lemma 2.16 by supposing that $\sum_{i=1}^{n} \alpha_i x_i = 0$. But then for each $j = i, \ldots, n$,

$$0 = \langle x_j, \underline{0}\rangle = \sum_{i=1}^{n} \alpha_i\langle x_j, x_i\rangle = \alpha_j\langle x_j, x_j\rangle.$$

But since $x_j \neq 0$, $\langle x_j, x_j\rangle > 0$ and so $\alpha_j = 0$ for each j. Thus X is a frame. Since the vectors in X are mutually orthogonal, X is an orthogonal basis for $\Re^n$.

For a symmetric matrix the roots of the characteristic equation will all be real. To see this in the 2×2 case, consider the characteristic equation

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - \lambda(a_{11} + a_{22}) = (a_{11}a_{22} - a_{21}a_{12}).$$

The roots of this equation are $\frac{-b+\sqrt{b^2-4c}}{2}$ with real roots iff $b^2 - 4c \geq 0$.

But this is equivalent to

$$(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{21}a_{12}) = (a_{11} - a_{22})^2 + 4(a_{12})^2 \geq 0,$$

since $a_{12} = a_{21}$.

Both terms in this expression are non-negative, and so λ_1, λ_2 are real.

In the case of a symmetric matrix, A, let E_λ be the set of eigenvectors associated with a particular eigenvalue, λ, of A together with the zero vector. Suppose x_1, x_2 belong to E_λ. Clearly $A(x_1 + x_2) = A(x_1) + A(x_2) = \lambda(x_1 + x_2)$ and so $x_1 +$

$x_2 \in E_\lambda$. If $x \in E_\lambda$, then $A(\alpha x) = \alpha A(x) = \alpha(\lambda x) = \lambda(\alpha x)$ and $\alpha x \in E_\lambda$ for each non-zero real number, α.

Since we also now suppose that for each eigenvalue, λ, the *eigenspace* E_λ contains $\underline{0}$, then E_λ will be a vector subspace of $\Re^n$. If $\lambda = \lambda_1 = \cdots = \lambda_r$ are repeated roots of the characteristic equation, then, in fact, the eigenspace, E_λ, will be of dimension r, and we can find r mutually orthogonal vectors in E_λ, forming a basis for E_λ.

Suppose now that A is a symmetric $n \times n$ matrix. As we shall show we may write $\wedge = P^{-1}AP$ where P is the $n \times n$ basis change matrix whose columns are the n linearly independent eigenvectors of A.

Now *normalise* each eigenvector x_j by defining $z_j = \frac{1}{\|x_j\|}(x_{1j}, \ldots, x_{nj})$ where $\|x_j\| = \sqrt{\sum (x_{kj})^2} = \sqrt{\langle x_j, x_j \rangle}$ is called the *norm* of x_j.

Let $Q = (z_1, \ldots, z_n)$ be the $n \times n$ matrix whose columns consist of $z_1, \ldots, z_n$. Now

$$Q^t Q = \begin{pmatrix} z_{11} & z_{21} & z_{n1} \\ z_{1j} & & z_{nj} \\ z_{1n} & & z_{nn} \end{pmatrix} \begin{pmatrix} z_{11} & z_{1j} & z_{1n} \\ z_{21} & z_{2j} & z_{2n} \\ \vdots & \vdots & \vdots \\ z_{n1} & z_{nj} & z_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \langle z_1, z_1 \rangle & \ldots & \langle z_1, z_n \rangle \\ \langle z_2, z_1 \rangle & & \\ \vdots & \ldots & \langle z_n, z_n \rangle \end{pmatrix}$$

since the (i, k)th entry in $Q^t Q$ is $\sum_{r=1}^n z_{ri} z_{rk} = \langle z_i, z_k \rangle$.

But $(z_i, z_k) = \langle \frac{x_i}{\|x_i\|}, \frac{x_k}{\|x_k\|} \rangle = \frac{1}{\|x_i\|\|x_k\|} \langle x_i, x_k \rangle = 0$ if $i \neq k$. On the other hand $\langle z_i, z_i \rangle = \frac{1}{\|x_i\|^2} \langle x_i, x_i \rangle = 1$, and $Q^t Q = I_n$, the $n \times n$ identity matrix. Thus $Q^t = Q^{-1}$.

Since $\{z_1, \ldots, z_n\}$ are eigenvectors of A with real eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ we obtain

$$\wedge = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_r & \\ 0 & & 0 \end{pmatrix} = Q^t A Q$$

where the last $(n - r)$ columns of Q correspond to the kernel vectors of A.

When A is a symmetric $n \times n$ matrix the function $A^* : \Re^n \times \Re^n \to \Re$ given by $A^*(x, y) = \langle x, A(y) \rangle$ is called a *quadratic form*, and in matrix notation is given by

$$(x_1, \ldots, x_n) \begin{pmatrix} a_{ij} \end{pmatrix} \begin{pmatrix} y_1 \\ \\ y_n \end{pmatrix}$$

Consider

$$\begin{aligned} A^*(x,x) &= \langle x, A(x)\rangle \\ &= \langle x, Q \wedge A^t(x)\rangle \\ &= \langle Q^t(x), \wedge\, Q^t(x)\rangle. \end{aligned}$$

Now $Q^t(x) = (x'_1 \dots, x'_n)$ is the coordinate representation of the vector x with respect to the new basis $\{z_1, \dots, z_n\}$ for $\Re^n$. Thus

$$A^*(x,x) = (x'_1, \dots, x'_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_r & \\ & & 0 \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_n \end{pmatrix} = \sum_{i=1}^{r} \lambda_i (x'_i)^2.$$

Suppose that rank $A = r$ and all eigenvalues of A are non-negative. In this case, $A^*(x,x) = \sum_{i=1}^{n} |\lambda_i|(x_i)^2 \geq 0$. Moreover if x is a non-zero vector then $Q^t(x) \neq 0$, since Q^t must have rank n.

Define the nullity of A^* to be $\{x : A^*(x,x) = 0\}$. Clearly if x is a non-zero vector in $\text{Ker}(A)$ then it is an eigenvector with eigenvalue 0. Thus the nullity of A^* is a vector subspace of $\Re^n$ of dimension at least $n-r$, where $r = \text{rank}(A)$. If the nullity of A^* is $\{\underline{0}\}$ then call A^* *non-degenerate*. If all eigenvalues of A are strictly positive (so that A^* is non-degenerate) then $A^*(x,x) > 0$ for all non-zero $x \in \Re^n$. In this case A^* is called *positive definite*. If all eigenvalues of A are non-negative but some are zero, then A^* is called *positive semi-definite*, and in this case $A^*(x,x) > 0$, for all x in a subspace of dimension r in $\Re^n$. Conversely if A^* is non-degenerate and all eigenvalues are strictly negative, then A^* is called *negative definite*. If the eigenvalues are non-positive, but some are zero, then A^* is called *negative semi-definite*.

The *index* of the quadratic form A^* is the maximal dimension of the subspace on which A^* is negative definite. Therefore index (A^*) is the number of strictly negative eigenvalues of A.

When A has some eigenvalues which are strictly positive and some which are strictly negative, then we call A^* a *saddle*.

We have not as yet shown that a symmetric $n \times n$ matrix has n real roots to its characteristic equation. We can show however that any (symmetric) quadratic form can be diagonalised.

Let $A = (a_{ij})$ and $\langle x, A(x)\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$. If $a_{ii} = 0$ for all $i = 1, \dots, n$ then it is possible to make a linear transformation of coordinates such that $a_{ij} \neq 0$ for some j. After relabelling coordinates we can take $a_{11} \neq 0$. In this case the quadratic form can be written

$$\begin{aligned} \langle x, A(x)\rangle &= a_{11}x_1^2 + 2a_{12}x_1x_2, \dots \\ &= a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 \cdots\right)^2 + \left(a_{22} - \frac{a_{12}^2}{a_{11}}^2\right)(x_2 + \cdots)^2 \;+ \cdots \\ &= \sum_{i=1}^{n} \alpha_i y_i^2. \end{aligned}$$

Here each y_i is a linear combination of $\{x_1, \dots, x_n\}$. Thus the transformation $x \to P(x) = y$ is non-singular and has inverse Q say.

Letting $x = Q(y)$ we see the quadratic form becomes

$$\begin{aligned}\langle x, A(x)\rangle &= \langle Q(y), A \circ Q(y)\rangle \\ &= \langle y, Q^t A Q(y)\rangle \\ &= \langle y, D(y)\rangle,\end{aligned}$$

where D is a diagonal matrix with real diagonal entries $(\alpha_1, \dots, \alpha_n)$. Note that $D = Q^t A Q$ and so $\operatorname{rank}(D) = \operatorname{rank}(A) = r$, say. Thus only r of the diagonal entries may be non zero. Since the symmetric matrix, A, can be diagonalised, not only are all its eigenvalues real, but its eigenvectors form a basis for $\Re^n$. Consequently $\wedge = P^{-1}AP$ where P is the $n \times n$ basis change matrix whose columns are these eigenvectors. Moreover, if λ is an eigenvalue with multiplicity r (*i.e.*, λ occurs as a root of the characteristic equation r times) then the eigenspace, E_λ, has dimension r. □

2.3.4 Examples

Example 2.7 To give an illustration of this procedure consider a matrix

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

representing the quadratic form $x_2^2 + 2x_1x_3$. Let

$$P_1(x) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

giving the quadratic form $z_1^2 - 2(z_2 - \frac{1}{2}z_3)^2 + \frac{1}{2}z_3^2$ and

$$P_2(z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Then $\langle x, A(x)\rangle = \langle y, D(y)\rangle$, where

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}, \quad \text{and} \quad A = P_1^t P_2^t D P_2 P_1.$$

Consequently the matrix A can be diagonalised. A has characteristic equation $(1-\lambda)(\lambda^2-1)$ with eigenvalues $1, 1, -1$.

Then normalized eigenvectors of A are

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix},$$

corresponding to the eigenvalues $1, 1, -1$.

Thus A^* is a non-degenerate saddle of index 1. Let Q be the basis change matrix

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0 & 1\\0 & 2 & 0\\1 & 0 & -1\end{pmatrix}.$$

Then

$$Q^t A Q = \frac{1}{2}\begin{pmatrix}1 & 0 & 1\\0 & 2 & 0\\1 & 0 & -1\end{pmatrix}\begin{pmatrix}0 & 0 & 1\\0 & 1 & 0\\1 & 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0 & 1\\0 & 2 & 0\\1 & 0 & -1\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & -1\end{pmatrix}.$$

As a quadratic form

$$(x_1, x_2, x_3)A\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = (x_1, x_2, x_3)\begin{pmatrix}x_3\\x_2\\x_1\end{pmatrix} = x_1x_3 + x_2^2.$$

We can also write this as

$$\begin{aligned}
&(x_1, x_2, x_3)\frac{1}{2}\begin{pmatrix}1 & 0 & 1\\0 & \sqrt{2} & 0\\1 & 0 & -1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & -1\end{pmatrix}\begin{pmatrix}1 & 0 & 1\\0 & \sqrt{2} & 0\\1 & 0 & -1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\\
&= \frac{1}{2}\left(x_1+x_3, \sqrt{2x_2}, x_1-x_3\right)\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & -1\end{pmatrix}\begin{pmatrix}x_1+x_3\\\sqrt{2x_2}\\x_1-x_3\end{pmatrix}\\
&= \frac{1}{2}(x_1+x_3)^2 + 2x_2^2 \ - (x_1-x_3)^2.
\end{aligned}$$

Note that A is positive definite on the subspace $\{(x_1, x_2, x_3) \in \Re^3 : (x_1 = x_3)\}$ spanned by the first two eigenvectors.

We can give a geometric interpretation of the behaviour of a matrix A with both positive and negative eigenvalues. For example

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

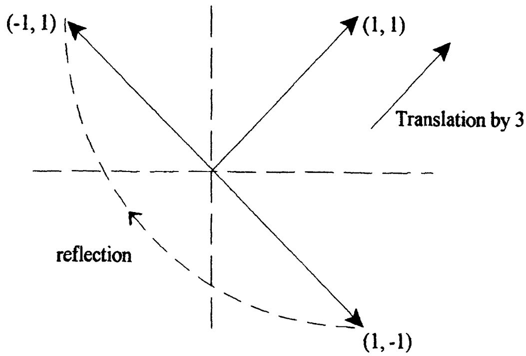


Fig. 2.2 A translation followed by a reflection

has eigenvectors

$$z_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad z_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

corresponding to the eigenvalues 3, -1 respectively. Thus A maps the vector z_1 to $3z_1$ and z_2 to $-z_2$. The second operation can be regarded as a reflection of the vector z_2 in the line $\{(x, y) : x - y = 0\}$, associated with the first eigenvalue. The first operation $z_1 \to 3z_1$ is a translation of z_1 to $3z_1$. Consider now any point $x \in \Re^2$. We can write $x = \alpha z_1 + \beta z_2$. Thus $A(x) = 3\alpha z_1 - \beta z_2$. In other words A may be decomposed into two operations: a translation in the direction z_1, followed by a reflection about z_1 as in Fig. 2.2.

2.4 Geometric Interpretation of a Linear Transformation

More generally suppose A has real roots to the characteristic equation and has eigenvectors $\{x_1, \ldots, x_s, z_1, \ldots, z_t, k_1, \ldots, k_p\}$.

The first s vectors correspond to positive eigenvalues, the next t vectors to negative eigenvalues, and the final p vectors belong to the kernel, with zero eigenvalues.

Then A may be described in the following way:

1. collapse the kernel vectors on to the image spanned by $\{x_1, \ldots, x_s, z_1, \ldots, z_t\}$.
2. translate each x_i to $\lambda_i x_i$.
3. reflect each z_j to $-z_j$, and then translate to $- \mid \mu_j \mid z_j$ (where μ_j is the negative eigenvalue associated with z_j).

These operations completely describe a symmetric matrix or a matrix, A, which is diagonalisable. When A is non-symmetric then it is possible for A to have complex roots.

For example consider the matrix

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

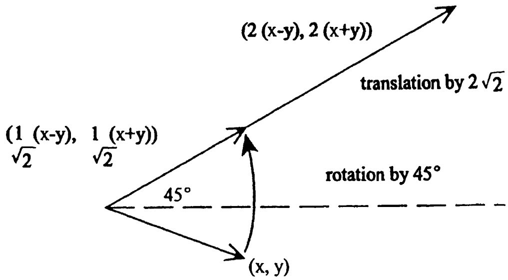


Fig. 2.3 A translation followed by a rotation

As we have seen this corresponds to a rotation by θ in an anticlockwise direction in the plane $\Re^2$. To determine the eigenvalues, the characteristic equation is $(\cos\theta - \lambda)^2 + \sin^2\theta = \lambda^2 - 2\lambda\cos\theta + (\cos^2\theta + \sin^2\theta) = 0$. But $\cos^2\theta + \sin^2\theta = 1$. Thus $\lambda = 2\cos\frac{\theta \pm 2\sqrt{\cos^2\theta - 1}}{2} = \cos\theta \pm\ i\sin\theta$.

More generally a 2×2 matrix with complex roots may be regarded as a transformation $\lambda e^{i\theta}$ where λ corresponds to a translation by λ and $e^{i\theta}$ corresponds to rotation by θ.

Example 2.8 Consider $A = \left(\begin{smallmatrix} 2 & -2 \\ 2 & 2 \end{smallmatrix}\right)$ with trace $(A) = \mathrm{tr}(A) = 4$ and $|A| = 8$.

As we have seen the characteristic equation for A is $(\lambda^2 - (\text{trace}\, A)) + |A| = 0$, with roots $\frac{\text{trace}(A) \pm \sqrt{(\text{trace}\, A)^2 - 4|A|}}{2}$. Thus the roots are $2 \pm \frac{1}{2}\sqrt{16 - 32} = 2 \pm 2i = 2\sqrt{2}[\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}]$ where $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$ and so $\theta = 45°$. Thus

$$A : \begin{pmatrix} x \\ y \end{pmatrix} \to 2\sqrt{2} \begin{pmatrix} x\cos 45 & -y\sin 45 \\ x\sin 45 & +y\cos 45 \end{pmatrix}.$$

Consequently A first sends (x, y) by a translation to $(2\sqrt{2}x, 2\sqrt{2}y)$ and then rotates this vector through an angle 45°.

More abstractly if A is an $n \times n$ matrix with two complex conjugate eigenvalues $(\cos\theta + i\sin\theta)$, $(\cos\theta - i\sin\theta)$, then there exists a two dimensional eigenspace E^θ such that $A(x) = \lambda e^{i\theta}(x)$ for all $x \in E_\theta$, where $\lambda e^{i\theta}(x)$ means rotate x by θ within E_θ and then translate by λ.

In some cases a linear transformation, A, can be given a *canonical form* in terms of rotations, translations and reflections, together with a collapse onto the kernel. What this means is that there exists a number of distinct *eigenspaces*

$$\{E_1, \ldots, E_p, X_1, \ldots, X_s, K\}$$

where A maps

1. E_j to E_j by rotating any vector in E_j through an angle θ_j;
2. X_j to X_j by translating a vector x in X_j to $\lambda_j x$, for some non-zero real number λ_j;

3. the kernel K to $\{0\}$.

In the case that the dimensions of these eigenspaces sum to n, then the canonical form of the matrix A is

$$\begin{pmatrix} e^{i\theta} & 0 & 0 \\ 0 & \wedge & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where $e^{i\theta}$ consists of p different 2×2 matrices, and $\wedge$ is a diagonal $s \times s$ matrix, while 0 is an $(n-r) \times (n-r)$ zero matrix, where $r = \text{rank}(A) = 2p + s$.

However, even when all the roots of the characteristic equation are real, it need not be possible to obtain a diagonal, canonical form of the matrix.

To illustrate, in Example 2.6 it is easy to show that the eigenvalue $\lambda = 0$ occurs twice as a root of the characteristic equation for the non-symmetric matrix A, even though the kernel is of dimension 1. The eigenvalue $\lambda = 7$ occurs once. Moreover the vector $(3, -29, 17)$ clearly must be an eigenvector for $\lambda = 7$, and thus span the image of A^2. However it is also clear that the vector $(3, -29, 17)$ does not span the image of A. Thus the eigenspace E_7 does not provide a basis for the image of A, and so the matrix A cannot be diagonalised.

However, as we have shown, for any *symmetric* matrix the dimensions of the eigenspaces sum to n, and the matrix can be expressed in canonical, diagonal, form.

In Chap. 4 below we consider smooth functions and show that "locally" such a function can be analysed in terms of the canonical form of a particular symmetric matrix, known as the Hessian.

Springer

http://www.springer.com/978-3-642-39817-9

Mathematical Methods in Economics and Social Choice
Schofield, N.
2014, XV, 262 p. 100 illus., Hardcover
ISBN: 978-3-642-39817-9