Module 2: Coordinates and Matrices

Remember: "*The introduction of numbers as coordinates is an act of violence.*" *H. Weyl*

The previous chapter introduced real vector spaces and their associated linear maps in an abstract setting. Here, we restrict attention to real, finite dimensional vector spaces.

1 Coordinates

Let V be a vector space of dimension n, and let $A = \{a_1, \ldots, a_n\}$ be a basis for V. Then $v \in V$ implies that we may write v uniquely as a linear combinations of elements of A, i.e.

$$v = \sum_{k=1}^{n} \alpha_k^v a_k.$$

The scalars $\{\alpha_1^v, \ldots, \alpha_n^v\}$ comprise the *coordinate representation* of v with respect to the basis A. Of course, if we chose a different basis for B then the vector v would have a different coordinate representation.

Here and in the sequel, it will be useful to refer to the Dirac function. In general, given two sets X and Y, $\delta : X \times Y \to \{0, 1\}$ is defined by $\delta_{xy} = 1$ if and only if $x = y$. The case $X = Y = \mathbb{Z}$ is illustrative: $\delta_{ij} = 1 \Leftrightarrow i = j$. I don't know why the argument of the Dirac function is typically written as a subscript with juxtaposition of the variables.

Next consider the n-dimensional vector space $\mathbb{R}^n$. As discussed in the previous chapter, the elements of $\mathbb{R}^n$ are *n-tuples*, which we will often write horizontally, i.e. $x \in \mathbb{R}^n$ implies $x = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}$; however you should always treat elements of $\mathbb{R}^n$ as *column vectors* unless explicitly stated otherwise – this will be critical for our results on matrices below. The *canonical basis* for $\mathbb{R}^n$ is $\mathscr{E} = \{e_1, \ldots, e_n\}$, where $e_i = (e_{1i}, \ldots, e_{ni})$ and $e_{ji} = \delta_{ji}$.

Now return to the abstract space V with basis A. Observe that given $v \in V$, the coordinate representation of v with respect to A defines an n-tuple $\alpha^v \in \mathbb{R}^n$. The relationship between V and the vector space of coordinate representations of elements of V with respect to a given basis A is provided in the following important exercise:

Exercise 1 *Define a linear map $\varphi : V \to \mathbb{R}^n$ by setting $\varphi(a_i) = e_i$ and extend linearly, as discussed in Chapter 2. Show that φ is an isomorphism.*

2 $\mathbb{R}^{m\times n}$

A real $m\times n$ *matrix* is a rectangular array of real numbers with m rows and n columns. Often you will hear someone say, "An n-dimensional vector is $n\times 1$ matrix." Of course, now you know what this statement really means: an element of a real, n-dimensional vector space has a coordinate representation with respect to a given basis that may be written as a column of n real numbers. I often think, and will sometimes say that an n-dimensional vector is $n\times 1$ matrix, but it is critical to keep in mind that this column of real numbers is only a representation of the vector, and the representation is basis-specific.

The collection of real $m\times n$ matrices may be suggestively written $\mathbb{R}^{m\times n}$. Matrices are typically denoted as follows: if $A\in\mathbb{R}^{m\times n}$ then $A=(a_{ij})$, where it is understood that $i=1,\ldots m$ and $j=1,\ldots,n$. Thus

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

a rectangular array of numbers.

Elements of $\mathbb{R}^{m\times n}$ may be added entry-wise, and scaled by elements of $\mathbb{R}$. Observing that the zero-matrix acts as an additive identity, we conclude that $\mathbb{R}^{m\times n}$ is a vector space. The canonical basis for $\mathbb{R}^{m\times n}$ is the set of matrices $\{e^{ij}\}$ with $(e^{ij}_{kl})=\delta_{(i,j)(k,l)}$.

Exercise 2 *Show that* $\mathrm{vec}:\mathbb{R}^{m\times n}\to\mathbb{R}^{mn}$*, defined on the basis* $\{e^{ij}\}$ *by*

$$e^{ij}\to e_{(j-1)m+i}$$

and extended linearly, is an isomorphism.

3 Linear functionals

Let V be a real vector space. Note that $\mathbb{R}$ may be viewed as a one-dimensional vector space. A *linear functional* is a linear map from V to $\mathbb{R}$. The set of linear functionals is called the *dual space* of V and is often denoted V^*. It is not difficult to show that V^* is a vector space under point-wise addition and scalar multiplication.

Now let V be a finite dimensional vector space with basis $A=\{a_1,\ldots,a_n\}$. Define $a_i^*:A\to\mathbb{R}$ on A by $a_i^*(a_j)=\delta_{ij}$, and then extend linearly to get a linear functional on V. Let $A^*=\{a_1^*,\ldots,a_n^*\}\subset V^*$. Now let $T\in V^*$. We claim that T is a linear combination of elements of A^*. We work as follows:

$$\begin{aligned}
T(v) &= T\left(\sum_{i=1}^{n} \alpha_i^v a_i\right) = \sum_{i=1}^{n} \alpha_i^v T(a_i) & (1)\\
&= \sum_{i=1}^{n} T(a_i)\left(\sum_{j=1}^{n} \alpha_j^v a_i^*(a_j)\right) & (2)\\
&= \sum_{i=1}^{n} T(a_i)a_i^*\left(\sum_{j=1}^{n} \alpha_j^v (a_j)\right) = \sum_{i=1}^{n} T(a_i)a_i^*(v). & (3)
\end{aligned}$$

To drive the point home, let $\beta_i = T(a_i)$. Then $v \in V$ implies $T(v) = \sum_{i=1}^{n} \beta_i a_i^*(v)$; or, more generally, $T = \sum_{i=1}^{n} \beta_i a_i^*$. Thus T is a linear combination of the elements of A^*. In fact, as you will demonstrate in the next exercise, A^* is a basis for V^*.

Exercise 3 *Show that A^* is a basis for V^*. This is the dual basis of the basis A.*

By exercise 3, if V be a finite dimensional vector space with basis $A = \{a_1, \ldots, a_n\}$ and $T \in V^*$ then the coordinate representation of T with respect to A^* is

$$(T(a_1), \ldots, T(a_n)) \in \mathbb{R}^n.$$

This suggests the following: define $\varphi_A : V^* \to V$ by $\varphi_A(T) = \sum_{i=1}^{n} T(a_i)a_i$.

Exercise 4 *Show that φ_A is an isomorphism.*

While the above development is somewhat abstract, careful inspection of (1) and (3) provides for very explicit computations. Fix the basis A of V. We see from (3) that the coordinate representation of T with respect to A^* is the element of $\mathbb{R}^n$ given by $(T(a_1), \ldots, T(a_n)) \in \mathbb{R}^n$; also, the coordinate representation of v is $\alpha^v \in \mathbb{R}^n$. Thus by (1), to evaluate $T(v)$ we simple add up the coordinate-wise products of coordinate representations of T and v: $T(v) = \sum_{i=1}^{n} \alpha_i^v T(a_i)$.

To put this on firmer ground, the previous exercise shows that $V^* \simeq \mathbb{R}^n \simeq V$. Thus any vector $v^* \in \mathbb{R}^n$ may be viewed as a linear functional taking $\mathbb{R}^n$ to $\mathbb{R}$. Write $v^* = (v_1^*, \ldots, v_n^*)$ and let $v \in \mathbb{R}$ be written $v = (v_1, \ldots, v_n)$. Then v^*, as a linear functional, acts on v, as a vector, as follows:

$$v^*(v) = \sum_{i=1}^{n} v_i^* v_i \equiv \langle v^*, v \rangle.$$

Exercise 5 *Recall that two vectors in $\mathbb{R}^n$ are* orthogonal *if $\langle v, w \rangle = 0$. Let $v \in \mathbb{R}^n$. Interpreted as a linear functional, what does the kernel of v look like? As an explicit example, let $n = 2$ and $v = (1, 2)$.*

A linear functional assigns a real number to a given vector so that addition and scalar multiplication are respected. In an important sense, a linear functional acts as an additive index, or summary statistic – concepts of particular use in Economics. The following exercise emphasizes this.

Exercise 6 *Show that the map take that takes a list of numbers to its mean is a linear functional.*

4 Matrices as linear maps

If P is an $m \times n$ matrix and Q is a $n \times k$ matrix then PQ is an $m \times k$ with the ij-entry of PQ being the inner product of the i^{th} row of P with the j^{th} column of Q:

$$pq_{ij} = \sum_{k=1}^{n} p_{ik} q_{kj}.$$

Note that matrix multiplication is *not* commutative. The reason for this mysterious definition of multiplication is revealed by viewing matrices as linear maps.

Let V and W be finite dimensional vector spaces with bases $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$, respectively. Let $T : V \to W$ be linear. Define

$$\beta(T)_j = (\beta(T)_{1j}, \ldots, \beta(T)_{nj}) \in \mathbb{R}^n$$

as follows:

$$T(a_j) = \sum_{i=1}^{m} \beta(T)_{ij} b_i. \tag{4}$$

Then for $v \in V$,

$$\begin{aligned} T(v) &= \left(\sum_{j=1}^{n} \alpha_j^v a_j \right) = \sum_{j=1}^{n} \alpha_j^v T(a_j) \\ &= \sum_{j=1}^{n} \alpha_j^v \left(\sum_{i=1}^{m} \beta(T)_{ij} b_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \beta(T)_{ij} \alpha_j^v \right) b_i. \end{aligned}$$

It follows that the i^{th} coordinate of the representation of $T(v)$ with respect to B is given by $\sum_{j=1}^{n} \beta(T)_{ij} \alpha_j^v$. Equivalently, we have the following diagram:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\varphi_A"'] & W \\ \mathbb{R}^n \arrow[r, "\beta(T)\alpha^v"'] & \mathbb{R}^m \arrow[u, "\varphi_B^{-1}"'] \end{tikzcd}$$

where φ_A and φ_B are the canonical isomorphisms and $\beta(T)\alpha^v$ is obtained by matrix multiplication.

In summary, matrices are *exactly* linear maps represented against bases. Specifically, given V and W finite dimensional vector spaces with bases $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_m\}$, a the matrix representing linear map $T : V \to W$ is $\beta(T)$ as given by 4. A simple way to remember the details is to observe that the matrix representation of T is the matrix whose columns are the coordinate-vectors of the images under T of the basis for V.

Exercise 7 *Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to \begin{pmatrix} x_1 - 4x_3 \\ x_2 \end{pmatrix}$$

Determine β^T. Find a basis for the kernel of T.

Exercise 8 *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^k$ be linear. Show that $\beta(S \circ T) = \beta(S)\beta(T)$. Thus matrix multiplication corresponds to composition of linear maps.*

4.1 The transpose

The *transpose* of an $m \times n$ matrix A is an $n \times m$ matrix A^T (sometimes written A' or, particularly in the complex case, A^*) with $a^T_{ij} = a_{ji}$, i.e. the rows and columns are switched (in the complex case, $a^*_{ij} = \overline{a}_{ji}$, i.e. each entry is conjugated – this is often referred to as the *Hermitian* transpose).

Exercise 9 *Show that if A and B are* conformable *matrices (i.e. there dimensions allow for the indicated product to exist) then $(AB)^T = B^T A^T$.*

If $v \in \mathbb{R}^n$ is viewed as a column matrix then v^T may be viewed as a *row* matrix, and its action as a linear functional (i.e. matrix multiplication) corresponds to an inner product: $v^T w = \langle v, w \rangle$. In this way, the transpose acts as a map from a vector space to its dual space.

Exercise 10 *Imagine a static, competitive consumer-choice environment such that a given agent must choose how much of each of n goods to consume, given an income level I in, say, dollars. To fix notation, let $\mathbb{R}_+ = [0, \infty)$, and $X = \mathbb{R}^n_+$ be* commodity space, *i.e. the set of all bundles of goods available to the consumer. Let $p \in \mathbb{R}^n_+$ be the vector of prices faced by the agent, i.e. a unit of good i costs p_i dollars.*

1. *As a linear functional, what does p^T measure?*
2. *Write an equation for the agent's budget constraint.*

3. *For $n = 2$ and 3, sketch a candidate budget constraint.*

Don't forget this exercise: prices are linear functionals – they are dual to quantities.

A final conceptual point is worth making. The transpose is really a *dual* concept. Specifically, if $S : V \to W$ is a linear map between vector spaces then we may define the dual map $S^* : W^* \to V^*$ as follows:

$$w^* \in W^* \text{ implies that } S^*(w^*) : V \to \mathbb{R} \text{ is given by } S^*(w^*)(v) = w^*(S(v)).$$

Letting $\varphi_V : V^* \to V$ and $\varphi_W : W^* \to W$ be the isomorphisms whose existence was established (with respect to an arbitrary basis) in exercise 5, the following diagram may be helpful:

$$\begin{tikzcd} V \arrow[r, "S"] & W \\ V^* \arrow[u, "\varphi_V"] & W^* \arrow[l, "S^*"] \arrow[u, "\varphi_W"'] \end{tikzcd}$$

If bases are chosen for V and W, and if $\beta(S)$ is the matrix representing S against these bases then $\beta(S)^T$ is the matrix representing S^* against the dual bases, i.e. $\beta(S^*) = \beta(S)^T$.

4.2 The determinant

The determinant, written det, is a map from $\mathbb{R}^{n\times n}$ to $\mathbb{R}$ with a number of magical properties, among them

1. $\det(I_n) = 1$
2. $\det(AB) = \det(A)\det(B)$
3. $\det(\alpha A) = \alpha^n \det(A)$

The formula for the determinant is ugly, and not worth presenting here. There are, however, two very important points worth emphasizing:

1. The determinant is the sum of products of entries in the matrix.[1] In particular – and don't forget this – the determinant may be viewed as a polynomial of degree n in its entries.

[1] Each product is signed in a manner that records whether the associated linear map preserves the orientation of the basis, but this is not important for our conceptual development.

2. Because it is a polynomial in its entries, the determinant is computable, involving only products and sums. The computation is tedious analytically and by hand, but trivial for a computer.

There are two natural ways to intuit the determinant. First, we consider the geometry of the matter. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear function (i.e. a square matrix), and let $\mathscr{S} \subset \mathbb{R}^n$ be the *unit cube*, that is, $\mathscr{S} = [0,1]^n$. Because T is linear, the image of $\mathscr{S}$ under the action of T is an m-dimensional parallelepiped for some $m \leq n$.

Exercise 11 *Graph the image of $\mathscr{S}$ in $\mathbb{R}^2$ under the linear maps given by the following three matrices:*

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

The determinant of T is the signed volume of this parallelepiped, where, again for intuition's sake, the sign is not important. In particular, if the dimension of the parallelepiped is less than n (as in the last matrix in the exercise above) then its volume as a subset of $\mathbb{R}^n$ is zero (just ask yourself what the volume of a square is as a subspace of $\mathbb{R}^3$), and thus the determinant of T is zero. As you might guess from the rank-nullity theorem, if the dimension of the parallelepiped is less than n then T is not preserving all information, i.e. the dimension of the kernel of T is not zero. In fact, we have the following extremely important property of the determinant:

Lemma 1 $\det T = 0 \Leftrightarrow \dim(ker(T)) > 0$.

The other way to intuit the determinant is via eigenvalues. Recall that $\lambda \in \mathbb{R}$ is an eigenvalue of T if there exists a non-zero vector $v \in \mathbb{R}^n$ such that $T(v) = \lambda v$. This may be written $(T - \lambda I_n)(v) = 0$, which means that $v \in \ker(T - \lambda I_n)$. It follows from Lemma 1 that $\det(T - \lambda I_n) = 0$.

Now remember the point I asked you not to forget: the determinant may be viewed as a polynomial of degree n in its entries. Thus, given T, we may form the function $\Phi_T(\lambda) = \det(T - \lambda I_n)$, which may be interpreted as a degree n polynomial in the indeterminate λ. The function $\Phi_T(\lambda)$ is called the *characteristic polynomial* of T, and the eigenvalues of T are *defined* to be the roots of this polynomial. We conclude, from the fundamental theorem of algebra, that every $n \times n$ matrix has exactly n eigenvalues, corresponding to the n roots of the characteristic polynomial; however, these roots, and thus these eigenvalues may be repeated, and, more importantly, may have imaginary components.[2]

[2] Note that in the later case, i.e. when λ is a non-real eigenvalue of a map $T : \mathbb{R}^n \to \mathbb{R}^n$, there is no vector $v \in \mathbb{R}^n$ so that $T(v) = \lambda v$: indeed, if $v \in \mathbb{R}^n$ then $\lambda v \notin \mathbb{R}^n$. This is why linear algebra is most naturally conducting over the complex numbers. To resolve this nuisance, we must embed V in $\mathbb{C}^n$, this is called *complexification*, and it is in its complexification that we will find the eigenvectors associated to the non-real eigenvalues.

Exercise 12 *Let p be a degree n complex polynomial. Show the following:*

1. *If $n \in \mathbb{N}$ and $z \in \mathbb{C}$ then $\overline{z^n} = \bar{z}^n$.*

2. $\overline{p(z)} = p(\bar{z})$.

3. *If $\lambda \in \mathbb{C}$ is a root of p then $\bar{\lambda}$ is as well.*

Thus non-real roots of complex polynomials come in conjugate pairs.

The determinant of T is equal to the product of the eigenvalues. It follows that $\det(T) = 0$ if and only if zero is an eigenvalue of T; or, said otherwise, if and only if there is a non-zero vector $v \in V$ so that $T(v) = 0$. Of course, this is just a confirmation of Lemma 1: the eigenspace associated to the zero eigenvalue is, after all, the kernel.

Exercise 13 *Use the previous exercise to show that the product of the eigenvalues of T is real.*

5 Invertibility

Let X and Y be sets and let $f : X \to Y$ be any map. We say f is *injective* provided that $f(x_1) = f(x_2)$ implies $x_1 = x_2$. Intuitively, injective functions don't lose information. We say f is *surjective* provided that for any $y \in Y$ there is an $x \in X$ so that $y = f(x)$. Intuitively, the image of X under the action of f is all of Y. A function with both properties is said to be *bijective*.

Exercise 14 *Show that $f : X \to Y$ is bijective if and only if it is invertible.*

Recall that a linear map $T : V \to W$ is invertible provided that there is a *linear* map $T^{-1} : W \to V$ so that $T^{-1} \circ T(v) = v$. By the previous exercise, we know that bijectivity is *necessary* for invertibility of T. In fact, it is sufficient as well.

Exercise 15 *Show that if the linear map $T : V \to W$ is bijective then it is invertible.*

Exercise 16 *Let $T : V \to W$ be linear. Show that the following are equivalent:*

1. *T is invertible*

2. $\dim(V) = \dim(W)$ *and* $\dim(\mathit{ker}(T)) = 0$

3. $\dim(V) = \dim(W) = \mathit{rank}(T)$

Now turn to matrix representations of linear maps. Naturally, a matrix is invertible provided that the associated linear map is invertible; and, similarly naturally, the matrix associated to the inverse of the linear map is called the inverse of the matrix. Thus if A is an invertible $(n \times n)$ matrix then there is a matrix A^{-1} so that $A^{-1}A = I_n$.

By item 2 of the previous exercise, the only possible invertible matrices are square and they must have trivial null-space. But now look at Lemma 1! We get the following amazing result, which is of the utmost practical importance:

Theorem 1 *A square matrix is invertible if and only if its determinant is non-zero.*

Why is this theorem so important? In economics, and in many other applied sciences, we are often faced with a system of nonlinear equations that might take the form $f(x,y) = 0$, where we need to solve for y (here, say, $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$). Now suppose we could *approximate* this system of nonlinear equations with a linear system of the form $Ay + Bx = 0$, where A is $n \times n$ and B is $n \times m$. Now remember what we emphasized above: the determinant is computable. Thus we can then compute $\det(A)$, and if it is non-zero, we now know there is a matrix A^{-1} so that $y = -A^{-1}Bx$: we solved for y. This is the miracle, and the reason linear maps are so important: their equations are the *only* equations we know how to solve.[3]

Exercise 17 *Let $A = \{a_1, \ldots, a_n\} \subset \mathbb{R}^n$. Let M be the $n \times n$ matrix whose columns are the elements of A. Show that A is a basis for $\mathbb{R}^n$ if and only if* $\det(M) \neq 0$.

6 Column and row space

Let $M \in \mathbb{R}^{m \times n}$, and denote by $\{M^i\}_{i=1}^m \subset \mathbb{R}^n$ the rows of M (but "written" as column vectors), and by $\{M_j\}_{j=1}^n \subset \mathbb{R}^m$ the columns of M. Then the *row space* of M is $\operatorname{span}\left(\{M^i\}_{i=1}^m\right)$ and the *column space* of M is $\operatorname{span}\left(\{M_j\}_{j=1}^n\right)$.

Exercise 18 *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map represented by the matrix M. Show that $T(V)$ is the column space of M.*

This exercise is very important: it says that the columns of M span the range of the associated linear map: intuitively, all of the information transferred by the linear map is expressed as linear combinations of the columns of M.

It turns out that the dimensions of the column and row spaces are equal. This seems perhaps unintuitive, unless you take a dual perspective. Recall that, given T, the dual T^* is defined

[3] Of course I have skipped one step: given $\det(A) \neq 0$, how does one compute A^{-1}? Again, the formulas are messy, but computers can handle the tedium.

as $T^*(w^*)(v) = w^*(T(v))$, where $w^* \in W^*$. It follows that $w^* \in \ker(T^*)$ if and only if $T(V) \subset \ker(w^*)$. Note that by the rank-nullity theorem, $\dim(\ker(w^*)) = m-1$.

Exercise 19 *Show that if $k \le m$ and $\{w_1^*, \dots, w_k^*\}$ are linearly independent then*

$$\dim\left(\bigcap_{j=1}^{k} \ker(w_j^*)\right) = m-k.$$

From this exercise, it follows that there are $m - \dim(T(V))$ linearly independent functionals in W^* satisfying $T(V) \subset \ker(w^*)$. Thus $\dim(\ker(T^*)) = m - \dim(T(V))$, and so, by the rank-nullity theorem,

$$\dim(T^*(W^*)) = \dim(T(V)). \tag{5}$$

But we know that $T(V)$ is the column space of M and, since the matrix representation of T^* is M^T, we know that $T^*(W^*)$ is the column space of M^T, which is the row space of M. We conclude

$$\begin{aligned} \text{dimension of rowspace of } M &= \text{dimension of columnspace of } M^T \\ &= \dim(T^*(W^*)) \\ &= \dim(T(V)) \\ &= \text{dimension of columnspace of } M \end{aligned}$$

where the third inequality comes from equation (5).

7 Matrix decompositions and Jordan form

7.1 Coordinate transforms

First we need to learn to change coordinates. Let V be a vector space with bases $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$. For $v \in V$, let α^v and β^v be the coordinate representations of v against A and B respectively. Next, define the β (no superscript) to be the $n \times n$ matrix with columns as the coordinate representations of the elements of A against the basis B, i.e. $a_j = \sum_i \beta_{ij} b_i$. Then

$$\begin{aligned} v &= \sum_j \alpha_j^v a_j = \sum_j \alpha_j^v \left(\sum_i \beta_{ij} b_i\right) \\ &= \sum_j \sum_i \alpha_j^v \beta_{ij} b_i = \sum_i \left(\sum_j \beta_{ij} \alpha_j^v\right) b_i. \end{aligned}$$

It follows that $\beta_i^v = \sum_j \beta_{ij}\alpha_j^v$, or $\beta^v = \beta\alpha^v$. We conclude that multiplying by the matrix β *changes coordinates* in the sense that it transforms an n-tuple representing a given vector against the basis A to an n-tuple representing the *same* vector against basis B.

Exercise 20 *Show that β^{-1} transforms coordinates from B to A.*

Now let $T : V \to V$ be a linear map, and let $M(T,A)$ be the matrix representation of T against the basis A. The following diagram shows that $M(T,B) = \beta M(T,A)\beta^{-1}$.

$$\begin{tikzcd}
V \arrow[r, "\varphi_A"] \arrow[d, "T"'] & \mathbb{R}^n \arrow[d, "M(T,A)"'] & \mathbb{R}^n \arrow[l, "\beta^{-1}"'] \arrow[d, "M(T,B)"] \\
V \arrow[r, "\varphi_B"'] & \mathbb{R}^n \arrow[r, "\beta"'] & \mathbb{R}^n
\end{tikzcd}$$

Here φ_A and φ_B are the canonical isomorphisms representing V against the bases A and B respectively.

In general, given matrices P and Q, if there is an invertible matrix S so that $Q = SPS^{-1}$ then we say that P and Q are *similar*. Thus two matrices are similar provided they represent the same linear transform against two (possibly) different bases.

Exercise 21 *Let $n = 2$ and T be a linear map represented by*

$$M = \begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}$$

against the canonical basis $\{(1,0),(0,1)\}$. Let $B = \{(2,1),(-2,1)\}$. Show that B is a basis for $\mathbb{R}^2$ and compute the matrix representation of M against it.

Changing coordinates can be very useful in applied work: it's like looking at the same problem from a different angle – sometimes all of the difficulties wash away. Perhaps the most common coordinate change involves representing matrices against a basis of eigenvectors. Specifically, let M is an $n \times n$ matrix representing a map $T : \mathbb{R}^n \to \mathbb{R}^n$ against the canonical basis. Suppose M has n linearly independent eigenvectors $\Xi = \{\xi_1, \ldots, \xi_n\}$ (and associated eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$). Let S be the matrix whose columns are the ξ_i. These columns may be interpreted as the representations of the eigenvectors against the canonical basis, and thus S may be viewed as the linear map that changes coordinates from representation against Ξ to represent representation against the canonical basis. Let Λ be the matrix representation of T against Ξ. Since we know that M is the representation of T against the canonical basis, our previous discussion shows $M = S\Lambda S^{-1}$, and thus $\Lambda = S^{-1}MS$. Now observe that MS is the matrix whose columns are $\lambda_i\xi_i$. It follows, as you are asked to demonstrate in the next exercise, that Λ is a diagonal matrix, and the diagonal elements correspond to the eigenvalues of M.

Exercise 22 *Show* $\Lambda_{ij} = \delta_{ij}\lambda_i$*.*

The product $S\Lambda S^{-1}$ is called the *eigenvalue decomposition* of M. It is ***extremely*** useful, as will be made evident later in this section. Intuitively, the eigenspaces provide the natural basis against which to measure a given linear map: if you know what happens on the eigenspaces, you know everything about the map; and, furthermore, you know *exactly* what happens on the eigenspaces: the map simply scales the vectors.

When an eigenvalue decomposition exists, the matrix is said to be *diagonalizable*, i.e. it is similar to a diagonal matrix. Not all matrices are diagonalizable. Indeed, suppose M is an $n \times n$ matrix whose entries are not all zeros, and suppose there is some $m \in \mathbb{N}$ such that $M^m = 0$ (i.e. the matrix will all zero entries). Then M cannot be diagonalizable. For suppose it was. Write $M = S\Lambda S^{-1}$, and note that since M is non-zero and S is invertible, it follows that $\Lambda = S^{-1}MS$ is not the zero matrix. Thus Λ has at least one non-zero entry, say Λ_{ii}. Now notice that $M^m = S\Lambda^n S^{-1}$, and since Λ is diagonal, $\Lambda^m_{ij} = \delta_{ij}\lambda_i^m$, where here Λ^m_{ij} is the ij-entry of Λ^m.[4] But $\lambda_i^m \neq 0$, so $\Lambda^m_{ii} \neq 0$, so $\Lambda^m \neq 0$, and thus $M^m \neq 0$, which is a contradiction.

Exercise 23 *Let M be the* 3×3 *matrix with zeros in every entry except that* $M_{12} = M_{13} = 1$*. Show that M is not diagonalizable.*

The following theorem provides a useful sufficient condition for diagonalizability.

Theorem 2 *If* $M \in \mathbb{R}^{n\times n}$ *has n distinct eigenvalues then M is diagonalizable.*

Exercise 24 *(hard) Show that matrices are generically diagonalizable.*

Two final comments: first, nobody said the entries of Λ are real; second, computers can very efficiently compute eigenvalues and eigenvectors.

7.2 Matrix structure and invariant subspaces

We need three definitions:

1. A matrix M is *upper triangular* if $M_{ij} = 0$ whenever $i > j$.
2. A complex matrix M is *unitary* if $M^*M = I_n = MM^*$, where M^* is the Hermitian (conjugate) transpose. A real matrix M is *orthogonal* if $M^TM = I_n = MM^T$.

[4]For a general square matrix M, he ij-entry of M^n is different from the ij-entry of M raised to the n.

3. A matrix M is *block diagonal* if it can be written as as diagonal matrix of square matrices, i.e.
$$M = \begin{pmatrix} M_1 & 0 & 0 & \cdots & 0 \\ 0 & M_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & M_m \end{pmatrix} \equiv \bigoplus_{i=1}^{m} M_i \tag{6}$$
with each M_i a square matrix, all 0s representing conformable matrices of zeros, and with the $\equiv$ defining notation, which will be examined in the following exercise.

Exercise 25 *Let $T : V \to V$ and $S : W \to W$ be linear maps of vectors spaces.*

1. *Define $T \oplus S : V \oplus W \to V \oplus W$ as follows: $T \oplus S(v,w) = (T(v), S(w))$. Show that $T \oplus S$ is a linear map. This map is called the direct sum of the maps T and S.*
2. *If $A = \{a_1, \ldots, a_n\}$ is a basis for V and $B = \{b_1, \ldots, b_m\}$ is a basis for W, show that*
$$A \oplus B = \{(a_1,0), \ldots, (a_n,0), (0,b_1), \ldots, (0,b_m)\}$$
is a basis for $V \oplus W$.
3. *Now let P and Q be matrix representations of T and S against A and B respectively. Show that $P \oplus Q$ is the matrix representation of $T \oplus S$ against the basis $A \oplus B$.*
4. *Show that $V \oplus 0 = \{(v,0) : v \in V\} \subset V \oplus W$ is a subspace of $V \oplus W$, and further, that $T \oplus S(V \oplus 0) \subset V \oplus 0$. Thus $V \oplus 0$ is an invariant subspace of $T \oplus S$.*

This exercise lays much of the groundwork for the following important simplification strategy: if an $n \times n$ matrix P is similar to a block diagonal matrix M of the form (6) then there are m subspaces V_i of $\mathbb{R}^n$ that are invariant under the action of M. Further, by changing coordinates, the action of P on these subspaces is decoupled, so each subspace can be examined in isolation:[5] more succinctly, to understand P we only need understand M; to understand M, we only need to understand the M_i. A common example is given in the next exercise.

Exercise 26 *Suppose $\dot{x} = Ax$ where $x \in \mathbb{R}^n$ an A is diagonalizable with real eigenvalues. Show that provided the eigenvalues are negative,* $\lim_{t\to\infty} x_i(t) = 0$.

We have learned that block-diagonal matrices, particularly diagonal matrices, are associated with very useful invariant subspace decompositions. An upper triangular matrix also provides for a useful characterization of the vector space in terms of invariant subspaces, but instead of a subspace structure that decouples the action of the map, an upper-triangular matrix provides for a nested sequence of subspaces which can be examined recursively.

[5] The precise notion of isolation is given by *orthogonality* which we will study soon.

Exercise 27 *Let $M \in \mathbb{R}^{n\times n}$ be upper triangular. Let $V_i \subset \mathbb{R}^n$ be the space spanned by $\{e_1,\dots,e_i\}$. Show that $v \in V_i \implies Mv \in V_i$.*

In this way, to understand M, we only need to understand the action of M on each of the subspaces V_i; and, given an understanding of the action of M on V_{i-1}, the determination of the action of M on V_i is simple in that it only requires understanding one new dimension.

7.3 Decompositions

Define a *Jordon* block J as follows: for given λ in $\mathbb{C}$ or $\mathbb{R}$, and $n \in \mathbb{N}$, let

$$J(\lambda,n) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}.$$

If $\lambda = a + bi$ with $b \neq 0$, define

$$C(\lambda) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Theorem 3 *Let $M \in \mathbb{R}^{n\times n}$ with eigenvalues $\{\lambda_1,\dots,\lambda_n\}$.*

1. *M is similar to a Jordon block diagonal matrix, i.e. a matrix of the form*

$$\bigoplus_{k=1}^{N} J(\lambda_{m_k},n_k) = \begin{pmatrix} J(\lambda_{m_1},n_1) & & & \\ & J(\lambda_{m_2},n_2) & & \\ & & \ddots & \\ & & & J(\lambda_{m_N},n_N) \end{pmatrix}.$$

   Note that there may be multiple Jordon blocks associated to the same eigenvalue, which is the reason for the strange m_k subscript on the λs.

2. *If λ is not real and $J(\lambda,m)$ is an associated Jordon block then there is necessarily a Jordon block of the form $J(\overline{\lambda},m)$. Furthermore, the direct sum $J(\lambda,m) \oplus J(\overline{\lambda},m)$ can be replaced with a $2m$ block of the form*

$$J^{\mathbb{R}}(\lambda,2m) = \begin{pmatrix} C(\lambda) & I_2 & & & \\ & C(\lambda) & I_2 & & \\ & & \ddots & \ddots & \\ & & & C(\lambda) & I_2 \\ & & & & C(\lambda) \end{pmatrix}.$$

Exercise 28 *What is the relationship between the Jordon decomposition and an eigenvalue decomposition?*

The Jordon decomposition, and its very useful special case, the eigenvalue decomposition, provide for the existence of a change of coordinates the decomposes the vector space into the smallest possible invariant subspaces, decomposes the matrix into a block diagonal that acts on each of these subspaces in isolation. While very useful, and fully general, it turns out that for computational purposes, a Schur decomposition is often preferable.

Theorem 4 *Let $M \in \mathbb{R}^{n\times n}$ with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$.*

1. *$M = QTQ^*$ where T is upper triangular with the eigenvalues of M on the diagonal, and Q is unitary.*
2. *$M = ZT^{\mathbb{R}}Z^T$ where $T^{\mathbb{R}}$ is upper block triangular and Z is orthogonal. The diagonal elements of $T^{\mathbb{R}}$ correspond to the eigenvalues of M. The real eigenvalues of M correspond to 1×1-blocks, and conjugate pairs of non-real eigenvalues of M $(\lambda, \bar{\lambda})$ correspond to 2×2 blocks of the form $C(\lambda)$.*

Thus a Schur decomposition provides for a nested sequence of invariant subspaces together with a matrix representation that acts on these subspaces recursively. A particularly nice feature of the Schur decomposition is that matrix inversion, which is computationally very expensive, is not needed – just take the transpose.

Exercise 29 *Let $A = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$. Define x_n recursively by $x_n = Ax_{n-1}$, where $x_n = (x_{1,n}, x_{2,n})$. Assume $x_{1,0} = 1$. Find $x_{2,0}$ so that $x_n \to 0$.*

8 Definiteness

The last topic in our introduction to matrices – and we could literally spend our lives discussing matrices – involves their behavior as bilinear forms. Fix a matrix $M \in \mathbb{R}^{n\times n}$ and observe that we may view M as a map from $\mathbb{R}^n \oplus \mathbb{R}^n$ to $\mathbb{R}$, by sending (v,w) to w^TMv. Note further that for fixed w, the map $v \to w^TMv$ is a linear functional, and for fixed v, the map $w \to w^TMv$ is a linear functional. For this reason, when acting in this fashion, M is called a *bilinear* form.

It is no coincidence that a linear functional (form) is a list with one index: $v = (v_i)$; and a bilinear form is a list with two indices: $M = (M_{ij})$. More generally, one may define a multilinear form as a list with a multi-index: these are called *tensors*, and are useful for higher-order approximations of non-linear maps. A vector is a first-order tensor, a matrix is a second-order tensor, and so on. The n^{th} derivative of a real valued function on $\mathbb{R}^n$ evaluated at a point is an n^{th}-order tensor.

Now some definitions:

1. A matrix M is *symmetric* if $M = M^T$.
2. A matrix is *positive definite* if $v^T M v > 0$ for all non-zero $v \in \mathbb{R}^n$.
3. A matrix is *positive semi-definite* if $v^T M v \geq 0$ for all non-zero $v \in \mathbb{R}^n$.

The analogous definitions hold for *negative* definiteness, etc. Note: Mas-Collel does not require a positive definite matrix to be symmetric; many authors do. So it will often be the case that if someone writes that M is positive definite, they are implicitly assuming M is symmetric.

Theorem 5 *If M is symmetric and positive semi-definite then the eigenvalues of M are real and non-negative.*

Exercise 30 *Show that if M and N are symmetric, positive definite then $M+N$ is too.*

Exercise 31 *Show that if M is symmetric, positive definite then there exists a square root.*

Positive (and negative) definiteness reflects concavity, which is the reason for its importance in economics. In particular, let $x = (x_1, \ldots, x_n)$ be n variables, and let $p(x)$ be a quadratic function, i.e. $p(x) = \sum_{ij} M_{ij} x_i x_j$. We see that $p(x) = x^T M x$. If we know that M is positive definite then we know that $x = 0$ is the unique minimum. Of course the functions we will want to minimize are not generally quadratic functions; but, as I alluded to above, the second derivative of a real function on $\mathbb{R}^n$ evaluated at a point is a second-order tensor, i.e. a matrix: if this matrix, which is necessarily symmetric, is positive semidefinite then we will be able to say the critical point identified by a first order condition is indeed the (local) minimum we seek.