

Lectures on Linear Algebra

Table of contents*	
Chapter 1 - in PDF**	Vector spaces - basics (has been reposted due to errors)
Chapter 2 - in PDF	Foundations of matrices
Chapter 3 - in PDF	Eigenvalues and Eigenvectors
Chapter 4 - in PDF	Unitary matrices
Chapter 5 - in PDF	Hermitian theory
Chapter 6 - in PDF	Normal matrices
Chapter 7 - in PDF	Factorization Theorems
Chapter 8 - in PDF	The Jordan Normal Form
Chapter 9 - in PDF	Hermitian and symmetric matrices*
Chapter 10 - in PDF	Positive matrices*

*Chapters 9 and 10 have the same link.

Math 640 - 600/700 — Linear Algebra for Applications

Fall 2003

Instructor: G. Donald Allen
Office: Milner 116A
Phone: 979-845-7950
Fax: 979-845-6028
dallen@math.tamu.edu
<http://www.math.tamu.edu/~dallen>

The place tamu is Texas A&M.

Chapter 1

Vectors and Vector Spaces

1.1 Vector Spaces

Underlying every vector space (to be defined shortly) is a scalar field F . Examples of scalar fields are the real and the complex numbers

$R :=$ real numbers

$\mathbb{C} :=$ complex numbers.

These are the only fields we use here.

Definition 1.1.1. A *vector space* V is a collection of objects with a (vector) addition and scalar multiplication defined that closed under both operations and which in addition satisfies the following axioms:

- (i) $(\alpha + \beta)x = \alpha x + \beta x$ for all $x \in V$ and $\alpha, \beta \in F$
- (ii) $\alpha(\beta x) = (\alpha\beta)x$
- (iii) $x + y = y + x$ for all $x, y \in V$
- (iv) $x + (y + z) = (x + y) + z$ for all $x, y, z \in V$
- (v) $\alpha(x + y) = \alpha x + \alpha y$
- (vi) $\exists 0 \in V \ni 0 + x = x$; 0 is usually called the *origin*
- (vii) $0x = 0$
- (viii) $ex = x$ where e is the multiplicative unit in F .

The “closed” property mentioned above means that for all $\alpha, \beta \in F$ and $x, y \in V$

$$\alpha x + \beta y \in V$$

(i.e. you can’t leave V using vector addition and scalar multiplication). Also, when we write for $\alpha, \beta \in F$ and $x \in V$

$$(\alpha + \beta)x$$

the ‘+’ is in the field, whereas when we write $x + y$ for $x, y \in V$, the ‘+’ is in the vector space. There is a multiple usage of this symbol.

Examples.

- (1) $R_2 = \{(a_1, a_2) \mid a_1, a_2 \in R\}$ two dimensional space.
- (2) $R_n = \{(a_1, a_2, \dots, a_n) \mid a_1, a_2, \dots, a_n \in R\}$, n dimensional space.
 $(a_1, a_2, \dots, a_n)$ is called an n -tuple.
- (3) $\mathbb{C}_2$ and $\mathbb{C}_n$ respectively to R_2 and R_n where the underlying field is $\mathbb{C}$, the complex numbers.
- (4) $P_n = \left\{ \sum_{j=0}^n a_j x^j \mid a_0, a_1, \dots, a_n \in R \right\}$ is called the *polynomial space* of all polynomials of degree n . Note this includes not just the polynomials of exactly degree n but also those of lesser degree.
- (5) $\ell_p = \{(a_i, \dots) \mid a_i \in R, \Sigma |a_i|^p < \infty\}$. This space is comprised of vectors in the form of infinite-tuples of numbers. Properly we would write

$$\ell_p(R) \quad \text{or} \quad \ell_p(\mathbb{C})$$

to designate the field.

- (6) $T_N = \left\{ \sum_{n=1}^N a_n \sin n\pi x \mid a_1, \dots, a_n \in R \right\}$, trigonometric polynomials.

Standard vectors in R^n

$$\begin{aligned} e_1 &= (1, 0, \dots, 0) \\ e_2 &= (0, 1, 0, \dots, 0) \\ e_3 &= (0, 0, 1, 0, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, \dots, 0, 1) \end{aligned}$$

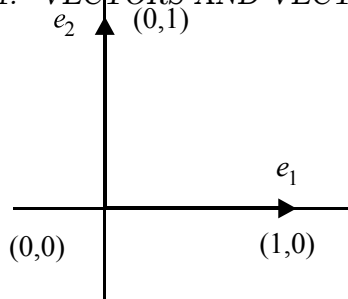
These are the *unit** vectors which point in the n *orthogonal** directions.

*Precise definitions will be given later.

For R^2 , the standard vectors are

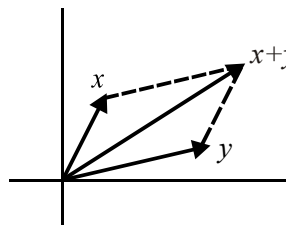
$$e_1 = (1, 0)$$

$$e_2 = (0, 1)$$



Graphical representation of e_1 and e_2 in the usual two dimensional plane.

Recall the usual vector addition in the plane uses the parallelogram rule

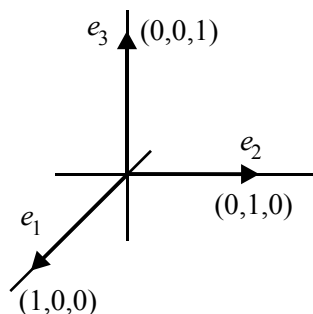


For R^3 , the standard vectors are

$$e_1 = (1, 0, 0)$$

$$e_2 = (0, 1, 0)$$

$$e_3 = (0, 0, 1)$$



Graphical representation of e_1 , e_2 , and e_3 in the usual

Linear algebra is the mathematics of vector spaces and their subspaces. We will see that many questions about vector spaces can be reformulated as questions about arrays of numbers.

1.1.1 Subspaces

Let V be a vector space and $U \subset V$. We will call U a *subspace* of V if U is closed under vector addition, scalar multiplication and satisfies all of the vector space axioms. We also use the term *linear subspace* synonymously.

Examples. Proofs will be given later

$$\begin{aligned} \text{let } V &= R^3 = \{(a, b, c) \mid a, b, c \in R\} \\ U &= \{(a, b, 0) \mid a, b \in R\}. \end{aligned} \quad (1.1)$$

Clearly $U \subset V$ and also U is a subspace of V .

$$\begin{aligned} \text{let } v_1, v_2 &\in R^3 \\ W &= \{av_1 + bv_2 \mid a, b \in R\} \\ W &\text{ is a subspace of } R^3. \end{aligned} \quad (1.2)$$

In this case we say W is “spanned” by $\{v_1, v_2\}$. In general, let $S \subset V$, a vector space, have the form

$$S = \{v_1, v_2, \dots, v_k\}.$$

The *span* of S is the set

$$U = \left\{ \sum_{j=1}^k a_j v_j \mid a_1, \dots, a_k \in R \right\}.$$

We will use the notion

$$\mathfrak{S}(v_1, v_2, \dots, v_k)$$

for the *span* of a set of vectors.

Definition 1.1.2. We say that

$$u = a_1 v_1 + \dots + a_k v_k$$

is a *linear combination* of the vectors $v_1, v_2, \dots, v_k$.

Theorem 1.1.1. *Let V be a vector space and $U \subset V$. If U is closed under vector addition and scalar multiplication, then U is a subspace of V .*

Proof. We remark that this result provides a “short cut” to proving that a particular subset of a vector space is in fact a subspace. The actual proof of this result is simple. To show (i), note that if $x \in U$ then $x \in V$ and so

$$(ab)x = ax + bx.$$

Now $ax, bx, ax + bx$ and $(a + b)x$ are all in U by the closure hypothesis. The equality is due to vector space properties of V . Thus (i) holds for U . Each of the other axioms is proved similarly. $\square$

A very important corollary follows about spans.

Corollary 1.1.1. *Let V be a vector space and $S = \{v_1, v_2, \dots, v_k\} \subset V$. Then $\mathfrak{S}(v_1, \dots, v_k)$ is a linear subspace of V .*

Proof. We merely observe that

$$\mathfrak{S}(v_1, \dots, v_k) = \left\{ \sum_1^k a_j v_j \mid a_1, \dots, a_k \in R \text{ or } C \right\}.$$

This means that the closure is built right into the definition of span. Thus, if

$$\begin{aligned} v &= a_1 v_1 + \dots + a_k v_k \\ w &= b_1 v_1 + \dots + b_k v_k \end{aligned}$$

then both

$$v + w = (a_1 + b_1)v_1 + \dots + (a_k + b_k)v_k$$

and

$$cv = ca_1 v + ca_2 v + \dots + ca_k v$$

are in U . Thus U is closed under both operations; therefore U is a subspace of V . $\square$

Example 1.1.1. (Product spaces.) Let V and W be vector spaces defined over the same field. We define the new vector space $Z = V \times W$ by

$$Z = \{(v, w) \mid v \in V, w \in W\}$$

We define vector addition as $(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$ and scalar multiplication by $\alpha(v, w) = (\alpha v, \alpha w)$. With these operations, Z is a vector space, sometimes called the *product* of V and W .

Example 1.1.2. Using set-builder notation, define $V_{13} = \{(a, 0, b) \mid a, b \in R\}$. Then U is a subspace of R_3 . It can also be realized as the subspace of the standard vectors $e_1 = (1, 0, 0)$ and $e_3 = (0, 0, 1)$, that is to say $V_{13} = \mathfrak{S}(e_1, e_3)$.

Example 1.1.3. More subspaces of R_3 . There are two other important methods to construct subspaces of R_3 . Besides the set builder notation used above, we have just considered the method of spanning sets. For example, let $S = \{v_1, v_2\} \subset R_3$. Then $\mathfrak{S}(S)$ is a subspace of R_3 . Similarly, if $T = \{v_1\} \subset R_3$. Then $\mathfrak{S}(T)$ is a subspace of R_3 . A third way to construct subspaces is by using inner products. Let $x, w \in R_3$. Expressed in coordinates $x = (x_1, x_2, x_3)$ and $w = (w_1, w_2, w_3)$. Define the inner product of x and w by $x \cdot w = x_1w_1 + x_2w_2 + x_3w_3$. Then $U_w = \{x \in R_3 \mid x \cdot w = 0\}$ is a subspace of R_3 . To prove this it is necessary to prove closure under vector addition and scalar multiplication. The latter is easy to see because the inner product is homogeneous in α , that is, $(\alpha x) \cdot w = \alpha x_1w_1 + \alpha x_2w_2 + \alpha x_3w_3 = \alpha(x \cdot w)$. Therefore if $x \cdot w = 0$ so also is $(\alpha x) \cdot w$. The additivity is also straightforward. Let $x, y \in U$. Then the sum

$$\begin{aligned}(x + y) \cdot w &= (x_1 + y_1)w_1 + (x_2 + y_2)w_2 + (x_3 + y_3)w_3 \\ &= (x_1w_1 + x_2w_2 + x_3w_3) + (y_1w_1 + y_2w_2 + y_3w_3) \\ &= 0 + 0 = 0\end{aligned}$$

However, by choosing two vectors $v, w \in R_3$ we can define $U_{v,w} = \{x \in R_3 \mid x \cdot v = 0 \text{ and } x \cdot w = 0\}$. Establishing $U_{v,w}$ is a subspace of R_3 is proved similarly. In fact, what is that both these sets of subspaces, those formed by spanning sets and those formed from the inner products are the same set of subspaces. For example, referring to the previous example, it follows that $V_{13} = \mathfrak{S}(e_1, e_3) = U_{e_2}$. Can you see how to correspond the others?

1.2 Linear independence and linear dependence

One of the most important problems in vector spaces is to determine if a given subspace is the span of a collection of vectors and if so, to determine a spanning set. Given the importance of spanning sets, we intend to examine the notion in more detail. In particular, we consider the concept of uniqueness of representation.

Let $S = \{v_1, \dots, v_k\} \subset V$, a vector space, and let $U = \mathfrak{S}(v_1, \dots, v_k)$ (or $\mathfrak{S}(S)$ for simpler notation). Certainly we know that any vector $v \in U$ has the representation

$$v = a_1v_1 + \dots + a_kv_k$$

for some set of scalars $a_1, \dots, a_k$. Is this representation unique? **Or**, can we

find another set of scalars $b_1, \dots, b_k$ not all the same as $a_1, \dots, a_k$ respectively for which

$$v = b_1 v_1 + \dots + b_k v_k.$$

We need more information about S to answer this question either way.

Definition 1.2.1. Let $S = \{v_1, \dots, v_k\} \subset V$, a vector space. We say that S is *linearly dependent* (l.d.) if there are scalars $a_1, \dots, a_k$ not all zero for which

$$a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0. \quad (\star)$$

Otherwise we say S is *linearly independent* (l.i.).

Note. If we allow all the scalars to be zero we can always arrange for $(\star)$ to hold, making the concept vacuous.

Proposition 1.2.1. *If $S = \{v_1, \dots, v_k\} \subset V$, a vector space, is linearly dependent, then one member of this set can be expressed as a linear combination of the others.*

Proof. We know that there are scalars $a_1, \dots, a_k$ such that

$$a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0$$

Since not all of the coefficients are zero, we can solve for one of the vectors as a linear combination of the other vectors. $\square$

Remark 1.2.1. Actually we have shown that there is *no* vector with a unique representation in $\mathfrak{S}(S)$.

Corollary 1.2.1. *If $0 \in S = \{v_1, \dots, v_k\}$, then S is linearly dependent.*

Proof. Trivial. $\square$

Corollary 1.2.2. *If $S = \{v_1, \dots, v_k\}$ is linearly independent then every subset of S is linearly independent.*

1.3 Bases

The idea of a basis is that of finding a minimal generating set for a vector space. Through basis, unicity of representation and a number of other useful properties, both theoretical and computational, can be concluded. Thinking of the concept in operations research ideas, a basis will be a redundancy free and complete generating set for a vector space

Definition 1.3.1. Let V be a vector space and $S = \{v_1, \dots, v_k\} \subset V$. We call S a *spanning set* for the subspace $U = \mathfrak{S}(S)$.

Suppose that V is a vector space, and $S = \{v_1, \dots, v_k\}$ is a linearly independent spanning set for V . Then S is called a *basis* of V . Modify this definition correspondingly for subspaces.

Proposition 1.3.1. *If S is a basis of V , then every vector has a unique representation.*

Proof. Let $S = \{v_1, \dots, v_k\}$ and $v \in V$. Then

$$v = a_1 v_1 + \dots + a_k v_k$$

for some choice of scalars. If there is a second choice of scalars $b_1, \dots, b_k$ not all the same, respectively, as $a_1, \dots, a_k$, we have

$$v = b_1 v_1 + \dots + b_k v_k$$

and

$$0 = (a_1 - b_1)v_1 + \dots + (a_k - b_k)v_k.$$

Since not all of the differences $a_1 - b_1, \dots, a_k - b_k$ are zero we must have that S is linearly dependent. This is a contradiction to our hypothesis, and the result is proved. $\square$

Example. Let $V = R^3$ and $S = \{e_1, e_2, e_3\}$. Then S is a basis for V .

Proof. Clearly V is spanned by S . Now suppose that

$$0 = a_1 e_1 + a_2 e_2 + a_3 e_3$$

or

$$\begin{aligned} (0, 0, 0) &= a_1(1, 0, 0) + a_2(0, 1, 0) + a_3(0, 0, 1) \\ &= (a_1, a_2, a_3). \end{aligned}$$

Hence $a_1 = a_2 = a_3 = 0$. Thus the set $\{e_1, e_2, e_3\}$ is linearly independent. $\square$

Remark 1.3.1. Note how we resolved the linearly dependent/linearly independent issue by converting a vector problem to a numbers problem. This is at the heart of linear algebra.

Exercise. Let $S = \{v_1, v_2\} = \{(1, 0, 1), (1, -1, 0)\} \subset \mathbb{R}^3$. Show that S is linearly independent and therefore a basis of $\mathfrak{S}(S)$.

1.4 Extension to a basis

In this section, we show that given a linearly independent set of vectors from a vector space with a finite spanning set, it is possible add to this set more vectors until it becomes a basis. Thus any set of linearly independent vectors can be a part (subset) of a basis.

Theorem 1.4.1 (Extension to a basis). *Assume that the given vector space V has a finite spanning set S_1 , i.e. $V = \mathfrak{S}(S_1)$. Let $S_0 = \{x_1, \dots, x_\ell\}$ be a linearly independent subset of V so that $\mathfrak{S}(S_0) \subsetneq V$. Then, there is a subset S'_1 of S_1 , such that $S_0 \cup S'_1$ is a basis for V .*

Proof. Our intention is to add vectors to S_0 keeping it linearly independent and eventually becoming a basis. There are a couple of steps.

Steps.

1. Since $\mathfrak{S}(S_1) \supsetneq \mathfrak{S}(S_0)$, there is a vector $y_1 \in S_1$ such that $S_{0,1} = \{S_0, y_1\}$ is linearly independent and thus $\mathfrak{S}(S_{0,1}) \supsetneq \mathfrak{S}(S_0)$.
2. Continue this process generating sets

$$\begin{aligned} S_{0,1} &= \{S_0, y_1\} \\ S_{0,2} &= \{S_{0,1}, y_2\} \\ &\vdots \\ S_{0,j} &= \{S_{0,j-1}, y_{j-1}\} \\ &\vdots \end{aligned}$$

At each step $S_{0,1}, S_{0,2}, \dots$ are linearly independent sets. Since S_1 is finite we must eventually have that

$$\mathfrak{S}(S_{0,m}) = \mathfrak{S}(S_1) = V.$$

3. Since $S_{0,m}$ is linearly independent and spans V , it must be a basis.

□

Remark 1.4.1. In the proof it was important to begin with *any* spanning set for V and to extract vectors from it as we did. Assuming merely that there exists a finite spanning set and extracting vectors directly from V leads to a problem of *terminus*. That is, when can we say that the new linearly independent set being generated in Step 2 above is a spanning set for V ? What we would need is a theorem that says something to the effect that if V has a finite basis, then every linearly independent set having the same number of vectors is also a basis. This result is the content of the next section. However, to prove it we need the Extension theorem.

Corollary 1.4.1. *If $S = \{v_1, \dots, v_k\}$ is linearly dependent then the representation of vectors in $\mathfrak{S}(S)$ is not unique.*

Proof. We know there are scalars $a_1, \dots, a_k$ not all zero, for which

$$a_1v_1 + \dots + a_kv_k = 0$$

let $v \in \mathfrak{S}(S)$ have the representation

$$v = b_1v_1 + b_2v_2 + \dots + b_kv_k.$$

Then we also have the representation

$$v = (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \dots + (a_k + b_k)v_k$$

establishing the result. □

Remark 1.4.2. The upshot of this construction is that we can *always* construct a basis from a spanning set. In actual practice this process may be quite difficult to carry out. In fact, we will spend some time achieving this goal. The main tool will be matrix theory.

1.5 Dimension

One of the most remarkable features of vector spaces is the notion of dimension. We need one simple result that makes this happen, the basis theorem.

Theorem 1.5.1 (Basis Theorem). *Let $S = \{v_1, \dots, v_k\} \subset V$ be a basis for V . Then every basis of V has k elements.*

Proof. We proceed by induction. Suppose $S = \{v_1\}$ and $T = \{w_1, w_2\}$ are both bases of V . Then since S is a basis

$$w_1 = \alpha_1 v_1 \quad w_2 = \alpha_2 v_1$$

and therefore

$$\frac{1}{\alpha_1} w_1 - \frac{1}{\alpha_2} w_2 = 0$$

which implies that T is linearly dependent (we tacitly assumed that both α_1 and α_2 were nonzero. Why can we do this?)

The next step is to assume the result holds for bases having up to k elements. Suppose that $S = \{v_1, \dots, v_{k+1}\}$ and $T = \{w_1, \dots, w_{k+2}\}$ are both bases of V . Now consider $S' = \{v_1, \dots, v_k\}$. We know that $\mathfrak{S}(S') \subsetneq \mathfrak{S}(S) = \mathfrak{S}(T) = V$. By our extension of bases result, there is a vector $w_{\ell_1} \in T$ such that

$$S'_1 = \{v_1, \dots, v_k, w_{\ell_1}\}$$

is linearly independent and $\mathfrak{S}(S'_1) \subset \mathfrak{S}(S) = V$. If $\mathfrak{S}(S'_1) \subsetneq V$, our extension result applies again to give a vector v_{ℓ_1} such that

$$S'_{11} = \{v_1, \dots, v_k, w_{\ell_1}, v_{\ell_1}\}$$

is linearly independent. The only possible selection is $v_{\ell_1} = v_{k+1}$. But in this case w_{ℓ_1} will depend on $v_1, \dots, v_k, v_{k+1}$, and that is a contradiction. Hence $\mathfrak{S}(v_1, \dots, v_k, w_{\ell_1}) = V$.

The next step is to remove the vector v_k from S'_1 and apply the extension to conclude that the span of the set

$$S'_2 = \{v_1, \dots, v_{k-1}, w_{\ell_1}, w_{\ell_2}\}$$

is V . We continue in this way eventually concluding that

$$S'_{k+1} = \{w_{\ell_1}, w_{\ell_2}, \dots, w_{\ell_{k+1}}\}$$

has span V . But $S'_{k+1} \subsetneq T$, whence T is linearly dependent. $\square$

Proposition 1.5.1 (Reduced spanning sets). (a) Suppose that $S = \{v_1, \dots, v_k\}$ spans V and v_j depends (linearly) on

$$S_j = \{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_k\}.$$

Then S_j also spans V .

- (b) If at least one vector in S is nonzero (that is $V \neq \{0\}$, the smallest vector space), then there is a subset $S_0 \subset S$ that is linearly independent and spans V .

Proof. (Left to reader.) □

Definition 1.5.1. The *dimension* of a vector space V is the (unique) number of vectors in a basis of V . We write $\dim(V)$ for the dimension.

Remark 1.5.1. This definition make sense possible only because of our basis theorem from which we are assured all every linearly independent spanning sets of V , that is all bases, have the same number of elements.

Examples.

$$(1) \dim(R_n) = n,$$

$$(2) \dim(P_n) = n + 1.$$

Exercise. Let M = all rectangular arrays of two rows and three columns with real entries. Find a basis for M , and find the dimension of M . Note

$$M = \left\{ \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \mid a, b, c, d, e, f \in R \right\}$$

Example 1.5.1. $P_n = \{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0\}$ is the vector space of polynomials of degree n . We claim that the powers, $x^0 = 1, x, x^2, \dots, x^n$ are linearly independent, and since

$$P_n = \mathfrak{S}(1, x, \dots, x^n)$$

they form a basis of P_n .

Proof. There are several ways we can prove this fact. Here is the most direct and it requires essentially no machinery. Suppose they are linearly dependent, which means that there are coefficients $a_0, a_1, \dots, a_n$ so that

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad (\star)$$

the function. (This functional view is critically important because every polynomial has roots.) There must be a coefficient *which is nonzero* and which corresponds to the highest power. Let us assume that $a_n \neq 0$, for convenience, and with no loss in generality.

Solve for x^n to get

$$x^n = -\frac{a_{n-1}}{a_n}x^{n-1} + \cdots + -\frac{a_1}{a_n}x - \frac{a_0}{a_n} \quad (\star\star)$$

Now compute the ratio of this expression divided by x^n on both sides, and let $x \rightarrow \infty$. The left side of course will be 1. Again for convenience we take $n = 2$. So, condensing terms we will have

$$\frac{b_1x + b_0}{x^2} = b_1 \left(\frac{1}{x} \right) + b_0 \left(\frac{1}{x^2} \right) = 1$$

where $b_j = -a_j/a_2$. But as $x \rightarrow \infty$ the expression $b_1 \left(\frac{1}{x} \right) + b_0 \left(\frac{1}{x^2} \right) \rightarrow 0$. This is a contradiction. It cannot be that the functions 1, x , and x^2 are linearly dependent.

In the general case for n we have

$$b_{n-1} \left(\frac{1}{x} \right) + b_{n-2} \left(\frac{1}{x^2} \right) + \cdots + b_0 \left(\frac{1}{x^n} \right) = 1,$$

where $b_j = -a_j/a_n$. Apply the same limiting argument to obtain the contradiction. Thus

$$T = \{1, x, \dots, x^n\}$$

is a basis of P_n .

A calculus proof is available. It is also based on the fact that if the powers are linearly independent and $(\star)$ holds, then we can assume that the same relation $(\star\star)$ is true. Now take the n^{th} derivative of both sides. We obtain

$$n! = 0$$

a contradiction, and the result is proved.

Finally, one more technique used to prove this result is by using the Fundamental Theorem of Algebra.

Theorem 1.5.2. *Every polynomial $(\star)$ of exactly n^{th} degree (i.e. with $a_n \neq 0$) has exactly n roots counted with multiplicity (i.e. if $q(x) = q_nx^n + q_{n-1}x^{n-1} + \cdots + q_1x + q_0 \in P_n(\mathbb{C})$, $q_n \neq 0$ then the number of solutions of $q(x) = 0$ is exactly n).*

From $(\star)$ above we have an n^{th} degree polynomial that is zero for every x . Thus the polynomial is zero, and this means all the coefficients are zero. This is a contradiction to the hypothesis, and therefore the theorem is proved. $\square$

Remark 1.5.2.

$$P_0 \subsetneq P_1 \subsetneq P_2 \subsetneq \cdots \subsetneq P_n \subsetneq \cdots .$$

On the other hand this is not true for the Euclidean spaces $R_1, R_2, \dots$. However, we may say that there is a subspace of R_3 which is “like” R_2 in every possible way. Do you see this? We have

$$\begin{aligned} R_2 &= \{(a, b) \mid a, b \in R\} \\ R_3 &= \{(a, b, c) \mid a, b, c \in R\}. \end{aligned}$$

No element in R_2 , an ordered *pair*, can be in R_3 , a set of ordered *triples*. However

$$U = \{(a, b, 0) \mid a, b \in R\}$$

is “like” R_2 in just about every way. Later on we will give a precise mathematical meaning to this comparison.

Example 1.5.2. Find a basis for the subspace V_0 of R_3 of all solutions to

$$x_1 + x_2 + x_3 = 0 \quad (\star)$$

where $x = (x_1, x_2, x_3) \in R^3$.

Solution. First show that the set $V_0 = \{(x_1, x_2, x_3) \in R_3 \mid x_1 + x_2 + x_3 = 0\}$ is in fact a subspace of R_3 . Clearly if $x = (x_1, x_2, x_3) \in V_0$ and $y = (y_1, y_2, y_3) \in V_0$ then $x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3) \in V_0$, proving closure under vector addition. Similarly V_0 is closed under scalar multiplication. Next, we seek a collection of vectors $v_1, v_2, \dots, v_k \in V_0$ so that $\mathfrak{S}(v_1, \dots, v_k) = V_0$. Let $x_3 = \alpha$ and $x_2 = \beta$ be free parameters. Then

$$x_1 = -(\alpha + \beta).$$

Hence all solutions of $(\star)$ have the form

$$\begin{aligned} x &= (-(\alpha + \beta), \beta, \alpha) \\ x &= \alpha(-1, 0, 1) + \beta(-1, 1, 0). \end{aligned}$$

Obviously the vectors $v_1 = (-1, 0, 1)$ and $v_2 = (-1, 1, 0)$ are linearly independent, and x is expressed as being in the span of them. So, $V_0 = \mathfrak{S}(v_1, v_2)$. V_0 has dimension 2.

Theorem 1.5.3 (Uniqueness). *Let $S = \{v_1, \dots, v_k\}$ be a basis of V . Then each vector $v \in V$ has a unique representation with respect to S .*

Proof. Since $\mathfrak{S}(S) = V$ we have that

$$v = a_1v_1 + a_2v_2 + \cdots + a_kv_k$$

for some coefficients $a_1, a_2, \dots, a_k$ in the given field. (This is the representation of v with respect to S .) If it is *not* unique there is another

$$v = b_1v_1 + b_2v_2 + \cdots + b_kv_k.$$

So, subtracting we have

$$(a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_k - b_k)v_k = 0$$

where the differences $a_j - b_j$ are not all zero. This implies that S is a linearly dependent set. $\square$

Theorem 1.5.4. *Suppose that $S = \{v_1, \dots, v_k\}$ is a basis of the vector space V . Suppose that $T = \{w_1, \dots, w_m\}$ is a linearly independent subset of V . Then $m \leq k$.*

Proof. We know that S is a linearly independent spanning set. This means that every linearly independent set of k vectors is also a spanning set. Therefore, $m > k$ renders a contradiction as $T_0 = \{w_1, \dots, w_k\}$ is a spanning set and $w_{k+1} \in \mathfrak{S}(T_0)$. $\square$

Definition 1.5.2. If A is any set we define

$$|A| := \text{cardinality of } A,$$

that is to say $|A|$ is the number of elements of A .

Example 1.5.3. Let $T = \{1, x, x^2, x^3\}$. Then $|T| = 4$.

Theorem 1.5.5. *Both R_k and $\mathbb{C}_k$ are k -dimensional and $S_k = \{e_1, e_2, \dots, e_k\}$ is a basis of both.*

Proof. It is easy to see that $e_1, \dots, e_k$ are linearly independent, and any vector x in R_k has the form

$$x = a_1e_1 + a_2e_2 + \cdots + a_ke_k$$

for $a_1, \dots, a_k \in R$. Thus S_k is a linearly independent spanning set and hence a basis of R_k . $\square$

Question: What single change to the proof above gives the theorem for C_k ?

The following results follow easily from previous results.

Theorem 1.5.6. *Let V be a k -dimensional vector space.*

- (i) *Every set T with $|T| > k$ is linearly dependent.*
- (ii) *If $D = \{v_1, \dots, v_j\}$ is linearly independent and $j < k$, then there are vectors $v_{\ell_1}, \dots, v_{\ell_{k-j}} \in V$ such that*

$$D \cup \{v_{\ell_1}, \dots, v_{\ell_{k-j}}\}$$

is a basis of V .

- (iii) *If $D \subset V$, $|D| = k$, and D is either a spanning set for V or linearly independent, then D is a basis for V .*

1.6 Norms

Norms are a way of putting a measure of distance on vector spaces. The purpose is for the refined analysis of vector spaces from the viewpoint of many applications. It is also to all the comparison of various vectors on the basis of their length. Ultimately, we wish to discuss vector spaces as representatives of points. Naturally, we are all accustomed to the “shortest distance” distance from the Pythagorean theorem. This is an example of a norm, but we shall consider them as real valued functions with very special properties.

Definition 1.6.1. *Norms on vector spaces over $\mathbb{C}$, or R . Let V be a vector space and suppose that $\|\cdot\|: V \rightarrow R^+$ is a function from V to the nonnegative reals for which*

- (i) $\|x\| \geq 0$ for all $x \in V$ and
 $\|x\| = 0$ if and only if $x = 0$
- (ii) $\|\alpha x\| = |\alpha| \|x\|$ for all $\alpha \in \mathbb{C}, R$ and $x \in V$
- (iii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ “The Triangle inequality”.

Then $\|\cdot\|$ is called a *norm* on V . The second condition is often termed the (positive) homogeneity property.

Remark 1.6.1. The notation is a substitute function notation. The expression $\|\cdot\|$, without the vector, is just the way a norm is expressed.

Examples. Let $V = R_n$ (or $\mathbb{C}_n$). Define for $x = (x_1, \dots, x_n)$

$$(i) \quad \|x\|_2 = (|x_1|^2 + |x_2|^2 + \dots + |x_n|^2)^{1/2} \quad \text{Euclidean norm}$$

$$(ii) \quad \|x\|_1 = (|x_1| + |x_2| + \dots + |x_n|) \quad \ell_1 \text{ norm}$$

$$(iii) \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad \ell_\infty \text{ norm}$$

Norm (ii) is read as: ell one norm. Norm (iii) is read as: ell infinity norm.

Proof that (ii) is a norm. Clearly (i) holds. Next

$$\begin{aligned} \|\alpha x\|_1 &= (|\alpha x_1| + |\alpha x_2| + \dots + |\alpha x_n|) \\ &= (|\alpha||x_1| + |\alpha||x_2| + \dots + |\alpha||x_n|) \\ &= |\alpha|(|x_1| + |x_2| + \dots + |x_n|) = |\alpha|\|x\|_1 \end{aligned}$$

which is what we needed to prove. Also,

$$\begin{aligned} \|x + y\|_1 &= (|x_1 + y_1| + |x_2 + y_2| + \dots + |x_n + y_n|) \\ &\leq (|x_1| + |y_1| + |x_2| + |y_2| + \dots + |x_n| + |y_n|) \\ &= (|x_1| + |x_2| + \dots + |x_n|) + (|y_1| + |y_2| + \dots + |y_n|) \\ &= \|x\|_1 + \|y\|_1. \end{aligned}$$

Here we used the fact that $|\alpha + \beta| \leq |\alpha| + |\beta|$ for numbers.

To prove that (i) is a norm we need a very famous inequality.

Lemma 1.6.1 (Cauchy–Schwartz). *Given that $a_1, \dots, a_n$ and $b_1, \dots, b_n$ are in $\mathbb{C}$. Then*

$$\sum_1^n |a_i b_i| \leq \left(\sum_1^n a_i^2 \right)^{1/2} \left(\sum_1^n b_i^2 \right)^{1/2}. \quad (\star)$$

Proof. We consider for the variable t

$$\Sigma(a_i + tb_i)^2 = \Sigma a_i^2 + 2t \Sigma a_i b_i + t^2 \Sigma b_i^2.$$

Note that $(\star)$ is obvious if $\sum_1^n a_i b_i = 0$. If not take

$$t = -\frac{\sum_1^n a_i^2}{\sum_1^n a_i b_i}.$$

Then

$$\begin{aligned}
 \Sigma(a_i + tb_i)^2 &= \Sigma a_i^2 - 2 \frac{\Sigma a_i^2}{\Sigma a_i b_i} \Sigma a_i b_i + \frac{(\Sigma a_i^2)^2}{(\Sigma a_i b_i)^2} \Sigma b_i^2 \\
 &= -\Sigma a_i^2 + \frac{(\Sigma a_i^2)^2 \Sigma b_i^2}{(\Sigma a_i b_i)^2} \\
 &= (\Sigma a_i^2) \left(-1 + \frac{\Sigma a_i^2 \Sigma b_i^2}{(\Sigma a_i b_i)^2} \right).
 \end{aligned}$$

Since the left side is ≥ 0 and since $\Sigma a_i^2 \geq 0$, we must have that

$$\left(-1 + \frac{\Sigma a_i^2 \Sigma b_i^2}{(\Sigma a_i b_i)^2} \right) \geq 0.$$

Solving this inequality we have

$$(\Sigma a_i b_i)^2 \leq \Sigma a_i^2 \Sigma b_i^2.$$

Now take square roots to get the result. $\square$

To prove that (i) is a norm, we note that conditions (i) and (ii) are straightforward. The truth of condition (iii) is a consequence of another famous result.

Theorem 1.6.1 (Minkowski). $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$.

Proof.

$$\begin{aligned}
 \Sigma(a_i + b_i)^2 &= \Sigma a_i^2 + 2 \Sigma a_i b_i + \Sigma b_i^2 \\
 &\leq \Sigma a_i^2 + 2 (\Sigma a_i^2)^{1/2} (\Sigma b_i^2)^{1/2} + \Sigma b_i^2 \\
 &= \left((\Sigma a_i^2)^{1/2} + (\Sigma b_i^2)^{1/2} \right)^2.
 \end{aligned}$$

Taking square roots gives the result. $\square$

Continuity and Equivalence of Norms

Lemma 1.6.2. *Every vector norm on $\mathbb{C}_n$ is continuous in the vector components.*

Proof. Let $x \in \mathbb{C}_n$ and $\|\cdot\|$ some norm on $\mathbb{C}_n$. We need to show that if the vector $\delta \rightarrow 0$, in components, then $\|x + \delta\| \rightarrow \|x\|$. First, by the triangle inequality

$$\begin{aligned} \|x + \delta\| &\leq \|x\| + \|\delta\| & \text{or} \\ \|x + \delta\| - \|x\| &\leq \|\delta\| \end{aligned}$$

Similarly

$$\begin{aligned} \|x\| &\leq \|x + \delta - \delta\| \\ &\leq \|x + \delta\| + \|\delta\| & \text{or} \\ -\|\delta\| &\leq \|x + \delta\| - \|x\| \end{aligned}$$

Therefore

$$|\|x + \delta\| - \|x\|| \leq \|\delta\|$$

Now expressing δ in components and standard bases vectors, we write $\delta = \delta_1 e_1 + \cdots + \delta_n e_n$ and

$$\begin{aligned} \|\delta\| &\leq |\delta_1| \|e_1\| + \cdots + |\delta_n| \|e_n\| \\ &\leq \max_{1 \leq i \leq n} |\delta_i| (\|e_1\| + \cdots + \|e_n\|) \\ &\leq M \max_{1 \leq i \leq n} |\delta_i| \end{aligned}$$

where $M = \|e_1\| + \cdots + \|e_n\|$. We know that if $\delta \rightarrow 0$ in components, then $\max_{1 \leq i \leq n} |\delta_i| \rightarrow 0$. Therefore $|\|x + \delta\| - \|x\|| \rightarrow 0$, as well. $\square$

Definition 1.6.2. Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be two vector norms on $\mathbb{C}_n$. We say that these norms are *equivalent* if there are positive constants m, M such that for all $x \in \mathbb{C}_n$

$$m \|x\|_a \leq \|x\|_b \leq M \|x\|_a$$

The remarkable fact about vector norms on $\mathbb{C}_n$ is that they are *all* equivalent. The only tool we need to prove this is the following result: *Every continuous function on a compact set of $\mathbb{C}_n$ assumes its maximum (and minimum) on that set.* The term “compact” refers to a particular kind of set K , one which is both bounded and closed. Bounded means that for $\max_{x \in K} \|x\| \leq B < \infty$ and closed means that if $\lim_{n \rightarrow \infty} x_n = x$, then $x \in K$.

Theorem 1.6.2. *All norms on $\mathbb{C}_n$ are equivalent.*

Proof. Since equivalence of norms is an equivalence condition, we can take one of the norms to be the infinity norm $\|\cdot\|_\infty$. Denote the other norm by $\|\cdot\|$. Now define $K = \{ x \mid \|x\|_\infty = 1 \}$. This set, called the unit ball in the infinity norm, is compact. Now we define

$$\begin{aligned} m &= \min_{x \in K} \|x\| \quad \text{and} \\ M &= \max_{x \in K} \|x\| \end{aligned}$$

Since $\|\cdot\|$ is a continuous function on K (from the lemma above) and since K is compact, we have that both the minimum and maximum are attained by specific vectors in K . Since these vectors are nonzero (they're in K) and because $\|x\|$ is positive for nonzero vectors, it must follow that $0 < m < M < \infty$. Hence, on K , the relation

$$m \|x\| \leq \|x\|_\infty \leq M \|x\|$$

holds true. For any vector $x \in \mathbb{C}_n$ we can write $x = \left(\frac{x}{\|x\|_\infty} \right) \|x\|_\infty$ and $\frac{x}{\|x\|_\infty} \in K$. Thus

$$\begin{aligned} m \left\| \frac{x}{\|x\|_\infty} \right\| &\leq \left\| \frac{x}{\|x\|_\infty} \right\|_\infty \leq M \left\| \frac{x}{\|x\|_\infty} \right\| \\ m \left\| \frac{x}{\|x\|_\infty} \right\| \|x\|_\infty &\leq \left\| \frac{x}{\|x\|_\infty} \right\|_\infty \|x\|_\infty \leq M \left\| \frac{x}{\|x\|_\infty} \right\| \|x\|_\infty \\ m \|x\| &\leq \|x\|_\infty \leq M \|x\| \end{aligned}$$

and the theorem is proved. $\square$

Example 1.6.1. Example. Find the estimates for the equivalence of $\|\cdot\|_2$ and $\|\cdot\|_\infty$

Solution. Let $x \in \mathbb{C}_n$. Then, because we know for any finite sequences

that $\sum_{i=1}^n |a_i b_i| \leq \max_{1 \leq i \leq n} |a_i| \sum_{i=1}^n |b_i|$

$$\begin{aligned}
 \|x\|_2 &= \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \\
 &\leq \left(\sum_{i=1}^n 1 \cdot |x_i|^2 \right)^{\frac{1}{2}} \\
 &\leq \left(\max_{1 \leq i \leq n} |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^n 1 \right)^{\frac{1}{2}} \\
 &= n^{\frac{1}{2}} \|x\|_\infty
 \end{aligned}$$

On the other hand, by the Cauchy-Schwartz inequality

$$\begin{aligned}
 \|x\|_\infty &= \max_{1 \leq i \leq n} |x_i| \\
 &\leq \sum_{i=1}^n |x_i| \\
 &\leq \left(\sum_{i=1}^n 1 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \\
 &= n^{\frac{1}{2}} \|x\|_2
 \end{aligned}$$

Putting these inequalities together we have

$$n^{-\frac{1}{2}} \|x\|_2 \leq \|x\|_\infty \leq n^{\frac{1}{2}} \|x\|_2$$

This makes $m = n^{-1/2}$ and $M = n^{\frac{1}{2}}$.

Remark 1.6.2. Note that the constants m and M depend on the dimension of the vector space. Though not the rule in all cases, it is mostly the situation.

Norms on polynomial spaces

Polynomial spaces, as we have considered earlier, can be given norms as well. Since they are function spaces, our norms usually need to consider all the values of the independent variable. In many, though not all, cases we need

to restrict the domains of the polynomials. With that in mind we introduce the notation

$$P_k(a, b) = P_k \text{ with domain restricted to the interval } [a, b]$$

We now define the function versions of the same three norms we have just studied. For functions $p(x)$ in $P_k(a, b)$ we define

1. $\|p(x)\|_2 = \left(\int_a^b |p(x)|^2 dx \right)^{\frac{1}{2}}$
2. $\|p(x)\|_1 = \int_a^b |p(x)| dx$
3. $\|p(x)\|_\infty = \max_{a \leq x \leq b} |p(x)|$

The positivity and homogeneity properties are fairly easy to prove. The triangle property is a little more involved. However, it has essentially been proved for the earlier norms. In the present case, one merely “integrates” over the inequality. Sometimes $\|\cdot\|_2$ is called the energy norm.

The integral norms are really the *norm* for polynomial spaces. Alternate norms use pointwise evaluation or even derivatives depending on the application. Here is a common type of norm that features the first derivative. For $p \in P_n(a, b)$ define

$$\|p\| = \max_{a \leq x \leq b} |p(x)| + \max_{a \leq x \leq b} |p'(x)|$$

As is evident this norm becomes large not only when the polynomial is large but also when its derivative is large. If we remove the term $\max_{a \leq x \leq b} |p(x)|$ from the norm above and define

$$N(p) = \max_{a \leq x \leq b} |p'(x)|$$

This function satisfies all the norm properties except one and thus is not a norm. (See the exercises.) Point evaluation-type norms take us too far afield of our goals partly because making point evaluations into norms requires some knowledge of interpolation and related topics. Leave it said that the obvious point evaluation functions such as $p(a)$ and the like will not provide us with norms.

1.7 Ordered Bases

Given a vector space V with a basis $S = \{v_1, v_2, \dots, v_k\}$ we now know that every vector $v \in V$ has a representation with respect to the basis

$$v = a_1v_1 + a_2v_2 + \dots + a_kv_k.$$

But no order is implied. For example, for R_2 we have $S = \{e_1, e_2\} = \{e_2, e_1\}$ shows us that there is no particular order conveyed through the definition of a basis. When we place an order on a basis we will notice an underlying algebraic structure of all k -dimensional vector spaces.

Definition 1.7.1. Let V be a k -dimensional vector space with basis $S = \{v_1; v_2; \dots; v_k\}$ is specified with a fixed and well defined order as indicated by their relevant positions. Then S is called an *ordered* basis. With ordered bases we obtain *coordinates*. Let V be a vector space of dimension k with ordered basis S , and suppose $v \in V$. For $1 \leq i \leq k$, we define the i^{th} *coordinate* of v with respect to S to be the i^{th} *coefficient* a_i in the representation

$$v = a_1v_1 + a_2v_2 + \dots + a_iv_i + \dots + a_kv_k.$$

In this way we can associate each $v \in V$ with a k -tuple of numbers $(a_1, a_2, \dots, a_k) \in R_k$ that are the coefficients of v with respect to S . The k -tuple is unique, owing to the fixed ordering of S . Conversely, for each ordered k -tuple $(a_1, a_2, \dots, a_k) \in R_k$ there is associated a unique vector $v \in V$ given by $v = a_1v_1 + a_2v_2 + \dots + a_iv_i + \dots + a_kv_k$.

We will express this association as

$$v \sim (a_1, a_2, \dots, a_k)$$

The following properties are each simple propositions:

- If $v \sim (a_1, a_2, \dots, a_k)$ and $w \sim (b_1, b_2, \dots, b_k)$ then

$$v + w \sim (a_1 + b_1, a_2 + b_2, \dots, a_k + b_k)$$

- If $v \sim (a_1, a_2, \dots, a_k)$ and $\alpha \in R$ (or $\mathbb{C}$), then

$$\alpha v \sim \alpha(a_1, a_2, \dots, a_k) = (\alpha a_1, \alpha a_2, \dots, \alpha a_k)$$

- If $v \sim (a_1, a_2, \dots, a_k) = (0, 0, \dots, 0)$, then $v = 0$.

We now define a special type of linear function from one linear space to another. The special condition is linearity of the map.

Definition 1.7.2. Let V and W be two vector spaces. We say that V and W are *homomorphic* if there is a mapping Φ between V and W for which

(1.) For v and w in V

$$\Phi(v + w) = \Phi(v) + \Phi(w)$$

(2.) For v and $\alpha \in R$ (or $\mathbb{C}$)

$$\Phi(\alpha v) = \alpha \Phi(v)$$

In this case we call Φ a *homomorphism* from V to W . Furthermore, we say that V and W are *isomorphic* if they are homomorphic and if

(3.) For each $w \in W$ there exists a unique $v \in V$ such that

$$\Phi(v) = w$$

In this case we call Φ a *isomorphism* from V to W .

We put all this together to show that finite dimensional vector spaces over the reals (resp. complex numbers) and the standard Euclidean spaces R_k (resp. $\mathbb{C}_k$) are very, very closely related. Indeed from the point of view of isometry, they are identical.

Theorem 1.7.1. *If V is a k -dimensional vector space over R (resp $\mathbb{C}$), then V is isomorphic to R_k (resp. $\mathbb{C}_k$).*

This constitutes the beginning of the sufficiency of matrix theory as a tool to study finite dimensional vector spaces.

Definition 1.7.3. The mapping $c_s : V \rightarrow R_k$ defined by

$$c_s(v) = (a_1, a_2, \dots, a_k)$$

where $v \sim (a_1, a_2, \dots, a_k)$ is the so-called *coordinate map*.

Example 1.7.1. We have shown that in R_3 the solutions to the equation $x_1 + x_2 + x_3 = 0$ for $x = (x_1, x_2, x_3) \in R_3$ is a subspace V_0 with basis $S = \{v_1, v_2\} = \{(-1, 0, 1), (-1, 1, 0)\}$. With respect to this basis $v_0 \in V_0$ if there are constants α_0, β_0 so that

$$v_0 = \alpha_0(-1, 0, 1) + \beta_0(-1, 1, 0)$$

With respect to this basis the coordinate map has the form

$$c_s(v_0) = (\alpha_0, \beta_0)$$

Therefore, we have established that V_0 is isomorphic to R_2 .

1.8 Exercises.

1. Show that $\{(a, b, 0) \mid a, b \in R\}$ is a subspace of R^3 by proving that it is spanned by vectors in R^3 . Find at least two sets of spanning sets.
2. Show that $\{(a, b, 1) \mid a, b \in R\}$ cannot be a subspace of R^3 .
3. Show that $\{(a - b, 2b - a, a - b) \mid a, b \in R\}$ is a subspace of R^3 by proving that it is spanned by vectors in R^3 .
4. For any $w \in R_3$, show that $U_w = \{x \in R_3 \mid x \cdot w = d \neq 0\}$ is **not** subspace of R_3 .
5. Find a set of vectors in R_3 that spans the subspace $U_w = \{x \in R_3 \mid x \cdot w = 0\}$, where $w = (1, 1, 1)$.
6. Why can $\{(a - b, a^2, ab) \mid a, b \in R\}$ never be a subspace of R^3 ?
7. Let $Q = \{x_1, \dots, x_k\}$ be a set of distinct points on the real line with $k < n$. Show that the subset P_Q of the polynomial space P_n of polynomials zero on the set Q is in fact a subspace of P_n . Characterize P_Q if $k > n$ and $k = n$.
8. In the product space defined above prove that definitions given the result is a vector space.
9. What is the product space $R_2 \times R_3$?
10. Find a basis for $Q = \{ax + bx^3 \mid a, b \in R\}$.
11. Let $T \subset P_n$ be those polynomials of exactly degree n . Show that T is not a subspace of P_n .
12. What is the dimension of

$$Q = \{ax + ax^2 + bx^3 \mid a, b \in Q\}.$$

What is a basis for Q ?

13. Given that $S = \{x_1, x_2, \dots, x_{2k}\}$ and $T = \{y_1, y_2, \dots, y_{2k}\}$ are both bases of a vector space V . (Note, the space V has dimension $2k$.) Consider the set of any k integers $L = \{l_1, \dots, l_k\} \subset \{1, 2, \dots, 2k\}$. (i) Show that associated with $P = \{x_{l_1}, x_{l_2}, \dots, x_{l_k}\}$ there are exactly k vectors from T , say $Q = \{y_{m_1}, y_{m_2}, \dots, y_{m_k}\}$ so that $P \cup Q$ is also a basis for V . (ii) Is the set of vectors from T unique? Why or why not?

14. Given P_n . Define $Z_n = \{p'(x) \mid p(x) \in P_n\}$. (The notation $p'(x)$ is the standard notation for the derivative of the function $p(x)$ with respect to the variable x .) What is another way to express Z_n in terms of previously defined spaces?
15. Show that $\|x\|_\infty$ is a norm. (Hint. The condition (iii) should be the focal point of your effort.)
16. Let $S = \{x_1, x_2, \dots, x_n\} \subset R_n$. For each $j = 1, 2, \dots, n$ suppose $x_j \in S$ has the property that its first $j - 1$ entries equal zero and the j^{th} entry is nonzero. Show that S is a basis of R_n .
17. Let $w = (w_1, w_2, w_3) \in R_3$, where all the components of w are strictly positive. Define $\|\cdot\|_w$ on R_3 by $\|x\|_w = \left(w_1 |x_1|^2 + w_2 |x_2|^2 + w_3 |x_3|^2\right)^{1/2}$. Show that $\|\cdot\|_w$ is a norm on R_3 .
18. Show that equivalence of norms is an equivalence relation.
19. Define for $p \in P_n(a, b)$ the function $N(p) = \max_{a \leq x \leq b} |p'(x)|$. Show that this is not a norm on $P_n(a, b)$.
20. For $p \in P_n(a, b)$ define $\|p\| = \max_{a \leq x \leq b} |p(x)| + \max_{a \leq x \leq b} |p''(x)|$. Show this is a norm on $P_n(a, b)$.
21. Suppose that V is a vector space with dimension k . Find two (linearly independent) spanning sets $S = \{v_1, v_2, \dots, v_k\}$ and $W = \{w_1, w_2, \dots, w_k\}$ of V such that if any $m < k$ vectors are chosen from S and any $k - m$ vectors are chosen from W , the resulting set will be a basis for V .
22. For $p \in P_n(a, b)$ define $N(p) = \left|p\left(\frac{a+b}{2}\right)\right|$. Show this is not a norm on $P_n(a, b)$.
23. Find the estimates for the equivalence of $\|\cdot\|_1$ and $\|\cdot\|_\infty$.
24. Find the estimates for the equivalence of $\|\cdot\|_1$ and $\|\cdot\|_2$.
25. Show that P_n defined over the reals is isomorphic to R_{n+1} .
26. Show that T_n , the space of trigonometric polynomials, defined over the reals is isomorphic to R_n .
27. Show that the product space $\mathbb{C}_k \times \mathbb{C}_m$ is isomorphic to $\mathbb{C}_{k+m}$.
28. What is the relation between the product space $P_n \times P_n$ and P_{2n} ? Find the polynomial space that is isomorphic to $P_n \times P_n$.

Terms.

Field

Vector space

scalar multiplication

Closed space

Origin

Polynomial space

Subspace, linear subspace

Span

Spanning set

Representation

Uniqueness of representation

linear dependence

linear independence

linear combination

Basis

Extension to a basis

Dimension

Norm

 ℓ_2 norm; ℓ_1 norm; $\ell_2\infty$ norm

Cauchy-Schwartz inequality

Fundamental Theorem of Algebra

Cardinality

Triangle inequality

Chapter 2

Matrices and Linear Algebra

2.1 Basics

Definition 2.1.1. A *matrix* is an $m \times n$ array of scalars from a given field F . The individual values in the matrix are called *entries*.

Examples.

$$A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

The *size* of the array is—written as $m \times n$, where

$$\begin{array}{ccc} & m \times n & \\ \nearrow & & \nwarrow \\ \text{number of rows} & & \text{number of columns} \end{array}$$

Notation

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{array}{l} \nwarrow \\ \longleftarrow \text{ rows} \\ \swarrow \end{array}$$

$$\begin{array}{c} \nwarrow \swarrow \nearrow \\ \text{columns} \end{array}$$

A := uppercase denotes a matrix

a := lower case denotes an entry of a matrix $a \in F$.

Special matrices

(1) If $m = n$, the matrix is called *square*. In this case we have

(1a) A matrix A is said to be *diagonal* if

$$a_{ij} = 0 \quad i \neq j.$$

(1b) A diagonal matrix A may be denoted by $\text{diag}(d_1, d_2, \dots, d_n)$ where

$$a_{ii} = d_i \quad a_{ij} = 0 \quad j \neq i.$$

The diagonal matrix $\text{diag}(1, 1, \dots, 1)$ is called the *identity* matrix and is usually denoted by

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

or simply I , when n is assumed to be known. $0 = \text{diag}(0, \dots, 0)$ is called the *zero matrix*.

(1c) A square matrix L is said to be *lower triangular* if

$$\ell_{ij} = 0 \quad i < j.$$

(1d) A square matrix U is said to be *upper triangular* if

$$u_{ij} = 0 \quad i > j.$$

(1e) A square matrix A is called *symmetric* if

$$a_{ij} = a_{ji}.$$

(1f) A square matrix A is called *Hermitian* if

$$a_{ij} = \bar{a}_{ji} \quad (\bar{z} := \text{complex conjugate of } z).$$

(1g) E_{ij} has a 1 in the (i, j) position and zeros in all other positions.

(2) A rectangular matrix A is called *nonnegative* if

$$a_{ij} \geq 0 \quad \text{all } i, j.$$

It is called *positive* if

$$a_{ij} > 0 \quad \text{all } i, j.$$

Each of these matrices has some special properties, which we will study during this course.

Definition 2.1.2. The set of all $m \times n$ matrices is denoted by $M_{m,n}(F)$, where F is the underlying field (usually R or $\mathbb{C}$). In the case where $m = n$ we write $M_n(F)$ to denote the matrices of size $n \times n$.

Theorem 2.1.1. $M_{m,n}$ is a vector space with basis given by E_{ij} , $1 \leq i \leq m$, $1 \leq j \leq n$.

Equality, Addition, Multiplication

Definition 2.1.3. Two matrices A and B are equal if and only if they have the same size and

$$a_{ij} = b_{ij} \quad \text{all } i, j.$$

Definition 2.1.4. If A is any matrix and $\alpha \in F$ then the scalar multiplication $B = \alpha A$ is defined by

$$b_{ij} = \alpha a_{ij} \quad \text{all } i, j.$$

Definition 2.1.5. If A and B are matrices of the same size then the *sum* A and B is defined by $C = A + B$, where

$$c_{ij} = a_{ij} + b_{ij} \quad \text{all } i, j$$

We can also compute the *difference* $D = A - B$ by summing A and $(-1)B$

$$D = A - B = A + (-1)B.$$

matrix subtraction.

Matrix addition “inherits” many properties from the field F .

Theorem 2.1.2. If $A, B, C \in M_{m,n}(F)$ and $\alpha, \beta \in F$, then

- (1) $A + B = B + A$ *commutativity*
- (2) $A + (B + C) = (A + B) + C$ *associativity*
- (3) $\alpha(A + B) = \alpha A + \alpha B$ *distributivity of a scalar*
- (4) If $B = 0$ (a matrix of all zeros) then

$$A + B = A + 0 = A$$

- (4) $(\alpha + \beta)A = \alpha A + \beta A$

$$(5) \alpha(\beta A) = \alpha\beta A$$

$$(6) 0A = 0$$

$$(7) \alpha 0 = 0.$$

Definition 2.1.6. If x and $y \in R_n$,

$$x = (x_1 \dots x_n)$$

$$y = (y_1 \dots y_n).$$

Then the scalar or dot product of x and y is given by

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

Remark 2.1.1. (i) Alternate notation for the scalar product: $\langle x, y \rangle = x \cdot y$.
(ii) The dot product is defined only for vectors of the same length.

Example 2.1.1. Let $x = (1, 0, 3, -1)$ and $y = (0, 2, -1, 2)$ then $\langle x, y \rangle = 1(0) + 0(2) + 3(-1) - 1(2) = -5$.

Definition 2.1.7. If A is $m \times n$ and B is $n \times p$. Let $r_i(A)$ denote the vector with entries given by the i^{th} row of A , and let $c_j(B)$ denote the vector with entries given by the j^{th} row of B . The product $C = AB$ is the $m \times p$ matrix defined by

$$c_{ij} = \langle r_i(A), c_j(B) \rangle$$

where $r_i(A)$ is the vector in R_n consisting of the i^{th} row of A and similarly $c_j(B)$ is the vector formed from the j^{th} column of B . Other notation for $C = AB$

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \begin{matrix} 1 \leq i \leq m \\ 1 \leq j \leq p. \end{matrix}$$

Example 2.1.2. Let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 \\ 3 & 0 \\ -1 & 1 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 1 & 2 \\ 11 & 4 \end{bmatrix}.$$

Properties of matrix multiplication

- (1) If AB exists, does it happen that BA exists and $AB = BA$? The answer is usually no. First AB and BA exist if and only if $A \in M_{m,n}(F)$ and $B \in M_{n,m}(F)$. Even if this is so the sizes of AB and BA are different (AB is $m \times m$ and BA is $n \times n$) unless $m = n$. However even if $m = n$ we may have $AB \neq BA$. See the examples below. They may be different sizes and if they are the same size (i.e. A and B are square) the entries may be different

$$\begin{array}{lll} A = [1, 2] & B = \begin{bmatrix} -1 \\ 1 \end{bmatrix} & AB = [1] \\ & & BA = \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix} \\ A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & B = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix} & AB = \begin{bmatrix} -1 & 3 \\ -3 & 7 \end{bmatrix} \\ & & BA = \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix} \end{array}$$

- (2) If A is square we define

$$\begin{aligned} A^1 &= A, & A^2 &= AA, & A^3 &= A^2A = AAA \\ A^n &= A^{n-1}A = A \cdots A \quad (n \text{ factors}). \end{aligned}$$

- (3) $I = \text{diag}(1, \dots, 1)$. If $A \in M_{m,n}(F)$ then

$$\begin{aligned} AI_n &= A \quad \text{and} \\ I_m A &= A. \end{aligned}$$

Theorem 2.1.3 (Matrix Multiplication Rules). Assume A, B , and C are matrices for which all products below make sense. Then

- (1) $A(BC) = (AB)C$
- (2) $A(B \pm C) = AB \pm AC$ and $(A \pm B)C = AC \pm BC$
- (3) $AI = A$ and $IA = A$
- (4) $c(AB) = (cA)B$
- (5) $A0 = 0$ and $0B = 0$

(6) For A square

$$A^r A^s = A^s A^r \quad \text{for all integers } r, s \geq 1.$$

Fact: If AC and BC are equal, it does not follow that $A = B$. See Exercise 60.

Remark 2.1.2. We use an alternate notation for matrix entries. For any matrix B denote the (i, j) -entry by $(B)_{ij}$.

Definition 2.1.8. Let $A \in M_{m,n}(F)$.

- (i) Define the *transpose* of A , denoted by A^T , to be the $n \times m$ matrix with entries

$$(A^T)_{ij} = a_{ji}.$$

- (ii) Define the *adjoint* of A , denoted by A^* , to be the $n \times m$ matrix with entries

$$(A^*)_{ij} = \bar{a}_{ji} \quad \text{complex conjugate}$$

Example 2.1.3.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 4 & 1 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & 5 \\ 2 & 4 \\ 3 & 1 \end{bmatrix}$$

In words... “The rows of A become the columns of A^T , taken in the same order.” The following results are easy to prove.

Theorem 2.1.4 (Laws of transposes). (1) $(A^T)^T = A$ and $(A^*)^* = A$

$$(2) (A \pm B)^T = A^T \pm B^T \quad (\text{and for } *)$$

$$(3) (cA)^T = cA^T \quad (cA)^* = \bar{c}A^*$$

$$(4) (AB)^T = B^T A^T$$

(5) If A is symmetric

$$A = A^T$$

(6) If A is Hermitian

$$A = A^*.$$

More facts about symmetry.

Proof. (1) We know $(A^T)_{ij} = a_{ji}$. So $((A^T)^T)_{ij} = a_{ij}$. Thus $(A^T)^T = A$.

(2) $(A \pm B)^T = a_{ji} \pm b_{ji}$. So $(A \pm B)^T = A^T \pm B^T$.

□

Proposition 2.1.1. (1) A is symmetric if and only if A^T is symmetric.

(1)* A is Hermitian if and only if A^* is Hermitian.

(2) If A is symmetric, then A^2 is also symmetric.

(3) If A is symmetric, then A^n is also symmetric for all n .

Definition 2.1.9. A matrix is called skew-symmetric if

$$A^T = -A.$$

Example 2.1.4. The matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ -2 & 3 & 0 \end{bmatrix}$$

is skew-symmetric.

Theorem 2.1.5. (1) If A is skew symmetric, then A is a square matrix and $a_{ii} = 0$, $i = 1, \dots, n$.

(2) For any matrix $A \in M_n(F)$

$$A - A^T$$

is skew-symmetric while $A + A^T$ is symmetric.

(3) Every matrix $A \in M_n(F)$ can be uniquely written as the sum of a skew-symmetric and symmetric matrix.

Proof. (1) If $A \in M_{m,n}(F)$, then $A^T \in M_{n,m}(F)$. So, if $A^T = -A$ we must have $m = n$. Also

$$a_{ii} = -a_{ii}$$

for $i = 1, \dots, n$. So $a_{ii} = 0$ for all i .

- (2) Since $(A - A^T)^T = A^T - A = -(A - A^T)$, it follows that $A - A^T$ is skew-symmetric.
- (3) Let $A = B + C$ be a second such decomposition. Subtraction gives

$$\frac{1}{2}(A + A^T) - B = C - \frac{1}{2}(A - A^T).$$

The left matrix is symmetric while the right matrix is skew-symmetric. Hence both are the zero matrix.

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

□

Examples. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is skew-symmetric. Let

$$\begin{aligned} B &= \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \\ B^T &= \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \\ B - B^T &= \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix} \\ B + B^T &= \begin{bmatrix} 2 & 1 \\ 1 & 8 \end{bmatrix}. \end{aligned}$$

Then

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

An important observation about matrix multiplication is related to ideas from vector spaces. Indeed, two very important vector spaces are associated with matrices.

Definition 2.1.10. Let $A \in M_{m,n}(\mathbb{C})$.

(i) Denote by

$$c_j(A) := j^{\text{th}} \text{ column of } A$$

$c_j(A) \in \mathbb{C}_m$. We call the subspace of $\mathbb{C}_m$ spanned by the columns of A the **column space** of A . With $c_1(A), \dots, c_n(A)$ denoting the columns of A

the column space is $\mathfrak{S}(c_1(A), \dots, c_n(A))$.

(ii) Similarly, we call the subspace of $\mathbb{C}_n$ spanned by the rows of A the **row space** of A . With $r_1(A), \dots, r_m(A)$ denoting the rows of A the row space is therefore $\mathfrak{S}(r_1(A), \dots, r_m(A))$.

Let $x \in \mathbb{C}_n$, which we view as the $n \times 1$ matrix $x = [x_1 \dots x_n]^T$. The product Ax is defined and

$$Ax = \sum_{j=1}^n x_j c_j(A).$$

That is to say, $Ax \in \mathfrak{S}(c_1(A), \dots, c_n(A)) = \text{column space of } A$.

Definition 2.1.11. Let $A \in M_n(F)$. The matrix A is said to be **invertible** if there is a matrix $B \in M_n(F)$ such that

$$AB = BA = I.$$

In this case B is called the *inverse* of A , and the notation for the inverse is A^{-1} .

Examples.

(i) Let

$$A = \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}$$

Then

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}.$$

(ii) For $n = 3$ we have

$$A = \begin{bmatrix} 1 & 2 & -1 \\ -1 & 3 & -1 \\ -2 & 3 & -1 \end{bmatrix} \quad A^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 3 & -2 \\ -3 & 7 & -5 \end{bmatrix}$$

A square matrix need not have an inverse, as will be discussed in the next section. As examples, the two matrices below do not have inverses

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 2 & 2 \end{bmatrix}$$

2.2 Linear Systems

The solutions of linear systems is likely the single largest application of matrix theory. Indeed, most reasonable problems of the sciences and economics that have the need to solve problems of several variable almost without exception are reduced to component parts where one of them is the solution of a linear system. Of course the entire solution process may have the linear system solver as a relatively small component, but an essential one. Even the solution of nonlinear problems, especially, employ linear systems to great and crucial advantage.

To be precise, we suppose that the coefficients a_{ij} , $1 \leq i \leq m$ and $1 \leq j \leq n$ and the data b_j , $1 \leq j \leq m$ are known. We define the *linear system* for the n unknowns $x_1, \dots, x_n$ to be

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \quad (*)$$

The *solution set* is defined to be the subset of R_n of vectors $(x_1, \dots, x_n)$ that satisfy each of the m equations of the system. The question of how to solve a linear system includes a vast literature of theoretical and computation methods. Certain systems form the model of what to do. In the systems below we note that the first one has three highly coupled (interrelated) variables.

$$\begin{aligned} 3x_1 - 2x_2 + 4x_3 &= 7 \\ x_1 - 6x_2 - 2x_3 &= 0 \\ -x_1 + 3x_2 + 6x_3 &= -2 \end{aligned}$$

The second system is more tractable because there appears even to the untrained eye a clear and direct method of solution.

$$\begin{aligned} 3x_1 - 2x_2 - x_3 &= 7 \\ x_2 - 2x_3 &= 1 \\ 2x_3 &= -2 \end{aligned}$$

Indeed, we can see right off that $x_3 = -1$. Substituting this value into the second equation we obtain $x_2 = 1 - 2 = -1$. Substituting both x_2 and x_3 into the first equation, we obtain $2x_1 - 2(-1) - (-1) = 7$, gives $x_1 = 2$. The

solution set is the vector $(2, -1, -1)$. The virtue of the second system is that the unknowns can be determined one-by-one, back substituting those already found into the next equation until all unknowns are determined. So if we can convert the given system of the first kind to one of the second kind, we can determine the solution.

This procedure for solving linear systems is therefore the applications of operations to effect the gradual elimination of unknowns from the equations until a new system results that can be solved by direct means. The operations allowed in this process must have precisely one important property: *They must not change the solution set by either adding to it or subtracting from it.* There are exactly three such operations needed to reduce any set of linear equations so that it can be solved directly.

(E1) Interchange two equations.

(E2) Multiply any equation by a nonzero constant.

(E3) Add a multiple of one equation to another.

This can be summarized in the following theorem

Theorem 2.2.1. *Given the linear system $(*)$. The set of equation operations E1, E2, and E3 on the equations of $(*)$ does not alter the solution set of the system $(*)$.*

We leave this result to the exercises. Our main intent is to convert these operations into corresponding operations for matrices. Before we do this we clarify which linear systems can have a solution. First, the system can be converted to matrix form by setting A equal to the $m \times n$ matrix of coefficients, b equal to the $m \times 1$ vector of data, and x equal to the $n \times 1$ vector of unknowns. Then the system $(*)$ can be written as

$$Ax = b$$

In this way we see that with $c_i(A)$ denoting the i^{th} column of A , the system is expressible as

$$x_1 c_1(A) + \cdots + x_n c_n(A) = b$$

From this equation it is clear that the system has a solution if and only if the vector b is in $\mathfrak{S}(c_1(A), \cdots, c_n(A))$. This is summarized in the following theorem.

Theorem 2.2.2. *A necessary and sufficient condition that $Ax = b$ has a solution is that $b \in \mathfrak{S}(c_1(A) \dots c_n(A))$.*

In the general matrix product $C = AB$, we note that the column space of $C \subset$ column space of A . In the following definition we regard the matrix A as a function acting upon vectors in one vector space with range in another vector space. This is entirely similar to the domain-range idea of function theory.

Definition 2.2.1. The **range** of $A = \{Ax \mid x \in R_n \text{ (or } \mathbb{C}_n \text{)}\}$.

It follows directly from our discussion above that the range of A equals $\mathfrak{S}(c_1(A), \dots, c_n(A))$.

Row operations: To solve $Ax = b$ we use a process called *Gaussian elimination*, which is based on row operations.

Type 1: Interchange two rows. (Notation: $R_i \longleftrightarrow R_j$)

Type 2: Multiply a row by a nonzero constant. (Notation: $cR_i \rightarrow R_i$)

Type 3: Add a multiple of one row to another row. (Notation: $cR_i + R_j \rightarrow R_j$)

Gaussian elimination is the process of reducing a matrix to its RREF using these row operations. Each of these operations is the respective analogue of the equation operations described above, and each can be realized by left matrix multiplication. We have the following.

Type 1

$$E_1 = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ \dots & \dots & 0 & \dots & \dots & \dots & 1 & \dots & \dots & \dots \\ & & & 1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & 1 & & & & \\ \dots & \dots & 1 & \dots & \dots & \dots & 0 & \dots & \dots & \dots \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix} \begin{array}{l} \text{row } i \\ \\ \text{row } j \end{array}$$

column i column j

Notation: $R_i \leftrightarrow R_j$

Type 2

$$E_2 = \begin{bmatrix} 1 & & & \vdots & & & \\ & \ddots & & \vdots & & & \\ & & 1 & \vdots & & & \\ \dots & \dots & \dots & c & \dots & \dots & \dots \\ & & & \vdots & 1 & & \\ & & & \vdots & & \ddots & \\ & & & \vdots & & & 1 \end{bmatrix} \quad \begin{array}{l} \text{row } i \\ \text{column } i \end{array}$$

Notation: cR_i

Type 3

$$E_3 = \begin{bmatrix} 1 & & \vdots & & & \\ & 1 & \vdots & & & \\ & & \vdots & & & \\ & & \vdots & \ddots & & \\ \dots & \dots & c & \dots & \dots & \dots \\ & & \vdots & & 1 & \\ & & \vdots & & & 1 \end{bmatrix} \quad \begin{array}{l} \text{row } j \\ \text{column } i \end{array}$$

Notation: $cR_i + R_j$, the abbreviated form of $cR_i + R_j \rightarrow R_j$

Example 2.2.1. The operations

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} \begin{array}{l} R_1 \longleftrightarrow R_2 \\ \rightarrow \end{array} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{array}{l} 4R_3 \\ \rightarrow \end{array} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -4 & 0 & 8 \end{bmatrix}$$

can also be realized as

$$\begin{aligned} R_1 &\longleftrightarrow R_2 : \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} \\ 4R_3 & : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -4 & 0 & 8 \end{bmatrix} \end{aligned}$$

The operations

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} \xrightarrow{\substack{-3R_1 + R_2 \\ 2R_1 + R_3}} \begin{bmatrix} 2 & 1 & 0 \\ -6 & -1 & 1 \\ 3 & 2 & 2 \end{bmatrix}$$

can be realized by the left matrix multiplications

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ -6 & -1 & 1 \\ 3 & 2 & 2 \end{bmatrix}$$

Note there are two matrix multiplications there, one for each Type 3 elementary operation.

Row-reduced echelon form. To each $A \in M_{m,n}(E)$ there is a canonical form also in $M_{m,n}(E)$ which may be obtained by row operations. Called the RREF, it has the following properties.

- (a) Each nonzero row has a 1 as the first nonzero entry ($:=$ *leading one*).
- (b) All column entries above and below a leading one are zero.
- (c) All zero rows are at the bottom.
- (d) The leading one of one row is to the left of leading ones of all lower rows.

Example 2.2.2.

$$B = \begin{bmatrix} 1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ is in RREF.}$$

Theorem 2.2.3. *Let $A \in M_{m,n}(F)$. Then the RREF is necessarily unique.*

We defer the proof of this result. Let $A \in M_{m,n}(F)$. Recall that the row space of A is the subspace of R_n (or $\mathbb{C}_n$) spanned by the rows of A . In symbols the row space is

$$\mathfrak{S}(r_1(A), \dots, r_m(A)).$$

Proposition 2.2.1. *For $A \in M_{m,n}(F)$ the rows of its RREF span the row space of A .*

Proof. First, we know the nonzero rows of the RREF are linearly independent. And all row operations are linear combinations of the rows. Therefore the row space generated from the RREF is contained in the row space of A . If the containment is proper. That is there is a row of A that is linearly independent from the row space of the RREF, this is a contradiction because every row of A can be obtained by the inverse row operations from the RREF. $\square$

Proposition 2.2.2. *If $A \in M_{m,n}(F)$ and a row operation is applied to A , then linearly dependent columns of A remain linearly dependent and linearly independent columns of A remain linearly independent.*

Proposition 2.2.3. *The number of linearly independent columns of $A \in M_{m,n}(F)$ is the same as the number of leading ones in the RREF of A .*

Proof. Let $S = \{i_1 \dots i_k\}$ be the columns of the RREF of A having a leading one. These columns of the RREF are linearly independent. Thus these columns were originally linearly independent. If another column is linearly independent, this column of the RREF is linearly dependent on the columns with a leading one. This is a contradiction to the above proposition. $\square$

Proof of Theorem 2.2.3. By the way the RREF is constructed, left-to-right, and top-to-bottom, it should be apparent that if the right most row of the RREF is removed, there results the RREF of the $m \times (n-1)$ matrix formed from A by deleting the n^{th} column. Similarly, if the bottom row of the RREF is removed there results a new matrix in RREF form, though not simply related to the original matrix A .

To prove that the RREF is unique, we proceed by a double induction, first on the number of columns. We take it as given that for an $m \times 1$ matrix the RREF is unique. It is either the zero $m \times 1$ matrix, which would be the case if A was zero or the matrix with a 1 in the first row and zeros in

the other rows. Assume therefore that the RREF is unique if the number of columns is less than n . Assume there are two RREF forms, B_1 and B_2 for A . Now the RREF of A is therefore unique through the $(n-1)^{\text{st}}$ columns. The only difference between the RREF's B_1 and B_2 must occur in the n^{th} column. Now proceed by induction on the number of nonzero rows. Assume that $A \neq 0$. If A has just one row, the RREF of A is simply the scalar multiple of A that makes the first nonzero column entry a one. Thus it is unique. If $A = 0$, the RREF is also zero. Assume now that the RREF is unique for matrices with less than m rows. By the comments above that the only difference between the RREF's B_1 and B_2 can occur at the (m, n) -entry. That is $(B_1)_{m,n} \neq (B_2)_{m,n}$. They are therefore not leading ones. (Why?) There is a leading one in the m^{th} row, however, because it is a non zero row. Because the row spaces of B_1 and B_2 are identical, this results in a contradiction, and therefore the (m, n) -entries must be equal. Finally, $B_1 = B_2$. This completes the induction. (Alternatively, the two systems pertaining to the RREF's must have the same solution set to the system $Ax = 0$. With $(B_1)_{m,n} \neq (B_2)_{m,n}$, it is easy to see that the solution sets to $B_1x = 0$ and $B_2x = 0$ must differ.) $\square$

Definition 2.2.2. Let $A \in M_{m,n}$ and $b \in R_m$ (or $\mathbb{C}_n$). Define

$$[A|b] = \begin{bmatrix} a_{11} & \dots & a_{1n} & b_1 \\ a_{21} & \dots & a_{2n} & b_2 \\ a_{m1} & \dots & a_{mn} & b_m \end{bmatrix}$$

$[A|b]$ is called the *augmented matrix* of A by b . $[A|b] \in M_{m,n+1}(F)$. The augmented matrix is a useful notation for finding the solution of systems using row operations.

Identical to other definitions for solutions of equations, the equivalence of two systems is defined via the idea of equality of the solution set.

Definition 2.2.3. Two linear systems $Ax = b$ and $Bx = c$ are called **equivalent** if one can be converted to the other by elementary equation operations.

It is easy to see that this implies the following

Theorem 2.2.4. Two linear systems $Ax = b$ and $Bx = c$ are equivalent if and only if both $[A|b]$ and $[B|c]$ have the same row reduced echelon form.

We leave the prove to the reader. (See Exercise 23.) Note that the solution set need not be a single vector; it can be null or infinite.

2.3 Rank

Definition 2.3.1. The **rank** of any matrix A , denote by $r(A)$, is the dimension of its column space.

Proposition 2.3.1. (i) *The rank of A equals the number of nonzero rows of the RREF of A , i.e. the number of leading ones.*

(ii) $r(A) = r(A^T)$.

Proof. (i) Follows from previous results.

(ii) The number of linearly independent rows equals the number of linearly independent columns. The number of linearly independent rows is the number of linearly independent columns of A^T —by definition. Hence $r(A) = r(A^T)$. $\square$

Proposition 2.3.2. *Let $A \in M_{m,n}(\mathbb{C})$ and $b \in \mathbb{C}_m$. Then $Ax = b$ has a solution if and only if $r(A) = r([A|b])$, where $[A|b]$ is the augmented matrix.*

Remark 2.3.1. Solutions may exist and may not. However, even if a solution exists, it may not be unique. Indeed if it is not unique, there is an infinity of solutions.

Definition 2.3.2. When $Ax = b$ has a solution we say the system is *consistent*.

Naturally, in practical applications we want our systems to be consistent. When they are not, this can be an indicator that something is wrong with the underlying physical model. In mathematics, we also want consistent systems; they are usually far more interesting and offer richer environments for study.

In addition to the column and row spaces, another space of great importance is the so-called null space, the set of vectors $x \in R_n$ for which $Ax = 0$. In contrast, when solving the simple single variable linear equation $ax = b$ with $a \neq 0$ we know there is always a unique solution $x = b/a$. In solving even the simplest higher dimensional systems, the picture is not as clear.

Definition 2.3.3. Let $A \in M_{m,n}(F)$. The **null space** of A is defined to be

$$\text{Null}(A) = \{x \in R_n \mid Ax = 0\}.$$

It is a simple consequence of the linearity of matrix multiplication that $\text{Null}(A)$ is a linear subspace of R_n . That is to say, $\text{Null}(A)$ is closed under vector addition and scalar multiplication. In fact, $A(x + y) = Ax + Ay = 0 + 0 = 0$, if $x, y \in \text{Null}(A)$. Also, $A(\alpha x) = \alpha Ax = 0$, if $x \in \text{Null}(A)$. We state this formally as

Theorem 2.3.1. *Let $A \in M_{m,n}(F)$. Then $\text{Null}(A)$ is a subspace of R_n while the range of A is in R_m .*

Having such solutions gives valuable information about the solution set of the linear system $Ax = b$. For, if we have found a solution, x , and have any vector $z \in \text{Null}(A)$, then $x + z$ is a solution of the same linear system. Indeed, what is easy to see is that if u and v are both solutions to $Ax = b$, then $A(u - v) = Au - Av = 0$, or what is the same $u - v \in \text{Null}(A)$. This means that to find all solutions to $Ax = b$, we need only find a single solution and the null space. We summarize this as the following theorem.

Theorem 2.3.2. *Let $A \in M_{m,n}(F)$ with null space $\text{Null}(A)$. Let x be any nonzero solution to $Ax = b$. Then the set $x + \text{Null}(A)$ is the entire solution set to $Ax = b$.*

Example 2.3.1. Find the null space of $A = \begin{bmatrix} 1 & 3 \\ -3 & -9 \end{bmatrix}$.

Solution. Solve $Ax = 0$. The RREF for A is $\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$. Solving $x_1 + 3x_2 = 0$, take $x_2 = t$, a “free” parameter and solve for x_1 to get $x_1 = -3t$. Thus every solution to $Ax = 0$ can be written in the form

$$x = \begin{bmatrix} -3t \\ t \end{bmatrix} = t \begin{bmatrix} -3 \\ 1 \end{bmatrix} \quad t \in R$$

Expressed this way we see that $\text{Null}(A) = \left\{ t \begin{bmatrix} -3 \\ 1 \end{bmatrix} \mid t \in R \right\}$, a subspace of R_2 of dimension 1.

Theorem 2.3.3 (Fundamental theorem on rank). *$A \in M_{m,n}(F)$. The following are equivalent*

- (a) $r(A) = k$.
- (b) *There exist exactly k linearly independent columns of A .*
- (c) *There exist exactly k linearly independent rows of A .*
- (d) *The dimension of the column space of A is k (i.e. $\dim(\text{Range } A) = k$).*
- (e) *There exists a set S of exactly k vectors in R_m for which $Ax = b$ has a solution for each $b \in \mathfrak{S}(S)$.*
- (f) *The null space of A has dimension $n - k$.*

Proof. The equivalence of (a), (b), (c) and (d) follow from previous considerations. To establish (e), let $S = \{c_{\ell_1}, c_{\ell_2}, \dots, c_{\ell_k}\}$ denote the linearly independent column vectors of A . Let $T = \{e_{\ell_1}, e_{\ell_2}, \dots, e_{\ell_k}\} \subset R_n$ be the standard vectors. Then $Ae_{\ell_j} = c_{\ell_j}$. If $b \in \mathfrak{S}(S)$, then $b = a_1c_{\ell_1} + a_2c_{\ell_2} + \dots + a_kc_{\ell_k}$. A solution to $Ax = b$ is given by $x = a_1e_{\ell_1} + a_2e_{\ell_2} + \dots + a_ke_{\ell_k}$. Conversely, if (e) holds, then the set S must be linearly independent for otherwise S could be reduced to $k - 1$ or fewer vectors. Similarly if A has $k + 1$ linearly independent columns then set S can be expanded. Therefore, the column space of A must have exactly k vectors.

To prove (f) we assume that $S = \{v_1, \dots, v_k\}$ is a basis for the column space of A . Let $T = \{w_1, \dots, w_k\} \subset R_n$ for which $Aw_i = v_i$, $i = 1, \dots, k$. By our extension theorem, we select $n - k$ vectors $w_{k+1}, \dots, w_n$ such that $U = \{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ is a basis of R_n . We must have that $Aw_{k+1} \in \mathfrak{S}(S)$. Hence there are scalars $b_1, \dots, b_k$ such that

$$Aw_{k+1} = A(b_1w_1 + \dots + b_kw_k)$$

and thus $w'_{k+1} = w_{k+1} - (b_1w_1 + \dots + b_kw_k)$ is in the null space of A . Repeat this process for each w_{k+j} , $j = 1, \dots, n - k$. We generate a total of $n - k$ vectors $\{w'_{k+1}, \dots, w'_n\}$ in this manner. This set must be linearly independent. (Why?) Therefore, the dimension of the null space must be at least $n - k$. Now we consider a new basis which consists of the original vectors and the $n - k$ vectors $\{w'_{k+1}, w'_{k+2}, \dots, w'_n\}$ for which $Aw = 0$. We assert that the dimension of the null space is exactly $n - k$. For if $z \in R_n$ is a vector for which $Az = 0$, then z can be uniquely written as a component z_1 from $\mathfrak{S}(T)$ and a component z_2 from $\mathfrak{S}(\{w'_{k+1}, \dots, w'_n\})$. But $Az_1 \neq 0$ and $Az_2 = 0$. Therefore $Az = 0$ is impossible unless the component $z_1 = 0$.

Conversely, if (f) holds we take a basis for the null space $T = \{u_1, u_2, \dots, u_{n-k}\}$ and extend the basis

$$T' = T \cup \{u_{n-k+1}, \dots, u_n\}$$

to R_n . Next argue similarly to above that

$$Au_{n-k+1}, Au_{n-k+2}, \dots, Au_n$$

must be linearly independent, for otherwise there is yet another linearly independent vector that can be added to its basis, a contradiction. Therefore the column space must have dimension at least, and hence equal to k . $\square$

The following corollary assembles many consequences of this theorem.

Corollary 2.3.1. (1) $r(A) \leq \min(m, n)$.

$$(2) \quad r(AB) \leq \min(r(A), r(B)).$$

$$(3) \quad r(A + B) \leq r(A) + r(B).$$

$$(4) \quad r(A) = r(A^T) = r(A^*) = r(\bar{A}).$$

(5) If $A \in M_m(F)$ and $B \in M_{m,n}(F)$, and if A is invertible, then

$$r(AB) = r(B).$$

Similarly, if $C \in M_n(F)$ is invertible and $B \in M_{m,n}(F)$

$$r(BC) = r(B).$$

$$(6) \quad r(A) = r(A^T A) = r(A^* A).$$

(7) Let $A \in M_{m,n}(F)$, with $r(A) = k$. Then $A = XBY$ where $X \in M_{m,k}$, $Y \in M_{k,n}$ and $B \in M_k$ is invertible.

(8) In particular, every rank 1 matrix has the form $A = xy^T$, where $x \in R_m$ and $y \in R_n$. Here

$$xy^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ \vdots & \vdots & & \vdots \\ x_m y_1 & x_m y_2 & \dots & x_m y_n \end{bmatrix}.$$

Proof. (1) The rank of any matrix is the number of linearly independent rows, which is the same as the number of linearly independent columns. The maximum this value can be is therefore the maximum of the minimum of the dimensions of the matrix, or $r(A) \leq \min(m, n)$.

(2) The product AB can be viewed in two ways. The first is as a set of linear combinations of the rows of B , and the other is as a set of linear combinations of the columns of A . In either case the number of linear independent rows (or columns as the case may be) In other words, the rank of the product AB cannot be greater than the number of linearly independent columns of A nor greater than the number of linearly independent rows of B . Another way to express this is as $r(AB) \leq \min(r(A), r(B))$

- (3) Now let $S = \{v_1, \dots, v_{r(A)}\}$ and $T = \{w_1, \dots, w_{r(B)}\}$ be basis of the column spaces of A and B respectively. Then, the dimension of the union $S \cup T = \{v_1, \dots, v_{r(A)}, w_1, \dots, w_{r(B)}\}$ cannot exceed $r(A) + r(B)$. Also, every vector in the column space of $A + B$ is clearly in the span of $S \cup T$. The result follows.
- (4) The rank of A is the number of linearly independent rows (and columns) of A , which in turn is the number of linearly independent columns of A^T , which in turn is the rank of A^T . That is, $r(A) = r(A^T)$. Similar proofs hold for A^* and $\bar{A}$.
- (5) Now suppose that $A \in M_m(F)$ is invertible and $B \in M_{m,n}$. As we have emphasized many times the rows of the product AB can be viewed as a set of linear combinations of the rows of B . Since A has rank m any set of linearly independent rows of B remains linearly independent. To see why, let $r_i(AB)$ denote the i^{th} row of the product AB . Then it is easy to see that

$$r_i(AB) = \sum_{j=1}^m a_{ij} r_j(B)$$

Suppose we can determine constants $c_1, \dots, c_m$ not all zero so that

$$\begin{aligned} 0 &= \sum_{j=1}^m c_j r_j(AB) = \sum_{i=1}^m c_i \sum_{j=1}^m a_{ij} r_j(B) \\ &= \sum_{j=1}^m r_j(B) \sum_{i=1}^m c_i a_{ij} \end{aligned}$$

This linear combination of the rows of B has coefficient given by $A^T c$, where $c = [c_1, \dots, c_m]^T$. Because the rank of A (and A^T) is m , we can solve this system for any vector $d \in R_m$. Suppose that the row vectors $r_{j_l}(B)$, $\ell = 1, \dots, r(B)$, are linearly independent. Arrange that the components of d to be zero for indices not included in the set j_l , $\ell = 1, \dots, r(B)$ and not all zero otherwise. Then the conclusion $0 = \sum_{j=1}^m r_j(B) \sum_{i=1}^m c_i a_{ij} = \sum_{l=1}^{r(B)} r_{j_l}(B) d_{j_l}$ is impossible. Indeed, the same basis of the row space of B will be a basis of the row space of AB . This proves the result.

- (6) We postpone the proof of this result until we discuss orthogonality.

- (7) Place A in RREF, say A_{RREF} . Since $r(A) = k$ we know the top k rows of A_{RREF} are linearly independent and the remaining rows are zero. Define Y to be the $k \times n$ matrix consisting of these top k rows. Define $B = I_k$. Now the rows of A are linear combinations of these rows. So, define the $m \times k$ matrix X to have rows as follows: The first row of consists of the coefficients so that $\sum x_{1j}r_j(Y) = r_1(A)$. In general, the i^{th} row of X is selected so that

$$\sum x_{ij}r_j(Y) = r_i(A)$$

- (8) This is an application of (7) noting in this special case that X is an $m \times 1$ matrix that can be interpreted as a vector $x \in R_m$. Similarly, Y is an $1 \times n$ matrix that can be interpreted as a vector $y \in R_n$. Thus, with $I = [1]$, we have

$$A = xy^T$$

□

Example 2.3.2. Here is the decomposition of the form given in Lemma 2.3.1 (7). The 3×4 matrix A has rank 2.

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ -1 & -2 & 3 \\ 2 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 3 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = XBY$$

The matrix Y is the RREF of A .

Example 2.3.3. Let $x = [x_1, x_2, \dots, x_m]^T \in R_m$ and $y = [y_1, y_2, \dots, y_n]^T \in R_n$. Then the rank one $m \times n$ matrix xy^T has the form

$$xy^T = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & & x_2y_n \\ \vdots & & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

In particular, with $x = [1, 3, 5]^T$, and $y = [-2, 7]^T$, the rank one 3×2 matrix xy^T is given by

$$xy^T = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} [-2, 7] = \begin{bmatrix} -2 & 7 \\ -6 & 21 \\ -10 & 35 \end{bmatrix}$$

Invertible Matrices

A subclass matrices $A \in M_n(F)$ that have only the zero kernel is very important in applications and theoretical developments.

Definition 2.3.4. $A \in M_n$ is called **nonsingular** if $Ax = 0$ implies that $x = 0$.

In many texts such matrices are introduced through an equivalent alternate definition involving rank.

Definition 2.3.5. $A \in M_n$ is **nonsingular** if $r(A) = n$.

We also say that nonsingular matrices have *full rank*. That nonsingular matrices are invertible and conversely together with many other equivalences is the content of the next theorem.

Theorem 2.3.4. *[Fundamental theorem on inverses] Let $A \in M_n(F)$. Then the following statements are equivalent.*

- (a) A is nonsingular.
- (b) A is invertible.
- (c) $r(A) = n$.
- (d) The rows and columns of A are linearly independent.
- (e) $\dim(\text{Range}(A)) = n$.
- (f) $\dim(\text{Null}(A)) = 0$.
- (g) $Ax = b$ is consistent for all $b \in R_n$ (or $\mathbb{C}_n$).
- (h) $Ax = b$ has a unique solution for every $x \in R_n$ (or $\mathbb{C}_n$).
- (i) $Ax = 0$ has only the zero solution.
- (j)* 0 is not an eigenvalue of A .
- (k)* $\det A \neq 0$.

* The statements about eigenvalues and the determinant ($\det A$) of a matrix will be clarified later after they have been properly defined. They are included now for completeness.

Definition 2.3.6. Two linear systems $Ax = b$ and $Bx = c$ are called **equivalent** if one can be converted to the other by elementary equation operations. Equivalently, the systems are equivalent if $[A \mid b]$ can be converted to $[B \mid c]$ by elementary row operations.

Alternatively, the systems are equivalent if they have the same solution set which means of course that both can be reduced to the same RREF.

Theorem 2.3.5. If $A \in M_n(F)$ and $B \in M_n(F)$ with $AB = I$, then B is unique.

Proof. If $AB = I$ then for every $e_1 \dots e_n$ there is a solution to the system $Ab_i = e_i$ for all $1 = 1, 2, \dots, n$. Thus the set $\{b_i\}_{i=1}^n$ is linearly independent (because the set $\{e_i\}$ is) and moreover a basis. Similarly if $AC = I$ then $A(C - B) = 0$, and there are $c_i \in R_n$ (or $\mathbb{C}_n$), $i = 1, \dots, n$ such that $Ac_i = e_i$. Suppose for example that $c_1 - b_1 \neq 0$. Since the $\{b_i\}_{i=1}^n$ is a basis it follows that $c_1 - b_1 = \sum \alpha_j b_j$, where not all α_j are zero. Therefore,

$$A(c_1 - b_1) = \sum \alpha_j Ab_j = \sum \alpha_j e_j \neq 0.$$

and this is a contradiction. $\square$

Theorem 2.3.6. Let $A \in M_n(F)$. If B is a right inverse, $AB = I$, then B is a left inverse.

Proof. Define $C = BA - I + B$, and assume $C \neq B$ or what is the same thing that B is not a left inverse. Then

$$\begin{aligned} AC &= ABA - A + AB = (AB)(A) - A + AB \\ &= A - A + AB = I \end{aligned}$$

This implies that C is another right inverse of A , contradicting Theorem 2.3.5. $\square$

2.4 Orthogonality

Let V be a vector space over $\mathbb{C}$. We define an inner product $\langle \cdot, \cdot \rangle$ on $V \times V$ to be a function from V to $\mathbb{C}$ that satisfies the following properties:

1. $\langle av, w \rangle = a \langle v, w \rangle$ and $\langle v, aw \rangle = \bar{a} \langle v, w \rangle$ ($\bar{a}$ is the complex conjugate of a)
2. $\langle v, w \rangle = \overline{\langle w, v \rangle}$

3. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ (linearity)
4. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$
5. $\langle v, v \rangle \geq 0$ with $\langle v, v \rangle = 0$ if and only if $v = 0$.

For inner products over real vector spaces, we neglect the complex conjugate operation. In addition, we want our inner products to define a *norm* as follows:

6. For any $v \in V$, $\|v\|^2 = \langle v, v \rangle$

We assume throughout the text that all vector spaces with inner products have norms defined exactly in this way. With the norm and vector v can be *normalized* by dilating it to have length 1, say $v_n = v \frac{1}{\|v\|}$. The simplest type of inner product on $\mathbb{C}_n$ is given by

$$\langle v, w \rangle = \sum_{i=1}^n x_i \bar{y}_i$$

We call this the **standard** inner product.

Using any inner product, we can define an angle between vectors.

Definition 2.4.1. The **angle** θ_{xy} between vectors x and y in R_n is defined by

$$\begin{aligned} \cos \theta_{xy} &= \frac{\langle x, y \rangle}{\|x\| \|y\|} \\ &= \frac{\langle x, y \rangle}{(\langle x, x \rangle)^{1/2} (\langle y, y \rangle)^{1/2}}. \end{aligned}$$

This comes from the well known result in R_2

$$x \cdot y = \|x\| \|y\| \cos \theta$$

which can be proved using the law of cosines. With angle comes the notion of orthogonality.

Definition 2.4.2. Two vectors u and v are said to be **orthogonal** if the angle between them is $\frac{\pi}{2}$ or what is the same thing $\langle u, v \rangle = 0$. In this case we commonly write $x \perp y$. We extend this notation to sets U writing $x \perp U$ to mean that $x \perp u$ for every $u \in U$. Similarly two sets U and V are called orthogonal if $u \perp v$ for every $u \in U$ and $v \in V$.

Remark 2.4.1. It is important to note that the notion of orthogonality depends completely on the inner product. For example, the **weighted inner product** defined by $\langle v, w \rangle = \sum_{i=1}^n w_i x_i \bar{y}_i$ where the $w_i > 0$ gives very different orthogonal vectors from the standard inner product.

Example 2.4.1. In R_n or $\mathbb{C}_n$ the standard unit vectors are orthogonal with respect to the standard inner product.

Example 2.4.2. In the R_3 the vectors $u = (1, 2, -1)$ and $v = (1, 1, 3)$ are orthogonal because

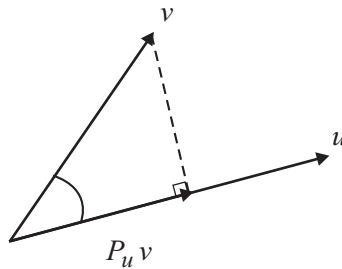
$$\langle x, y \rangle = 1(1) + 2(2) - 1(3) = 0$$

Note that in R_3 the complex conjugate is not written. The set of vectors $(x_1, x_2, x_3) \in R_3$ orthogonal to $u = (1, 2, -1)$ satisfies the equation $x_1 + 2x_2 - x_3 = 0$ is recognizable as the plane with normal vector u .

Definition 2.4.3. We define the *projection* $P_u v$ of one vector v in the direction of an other vector u to be

$$P_u v = \frac{\langle u, v \rangle}{\|u\|^2} u$$

As you can see, we have merely written an expression for the more intuitive version of the projection in question given by $\|v\| \cos \theta_{uv} \frac{u}{\|u\|}$. In the figure below, we show the fundamental diagram for the projection of one vector in the direction of another.



If the vectors u and v are orthogonal, it is easy to see that $P_u v = 0$. (Why?)

Example 2.4.3. Find the projection of the vector $v = (1, 2, 1)$ on the vector $u = (-2, 1, 3)$

Solution. We have

$$\begin{aligned} P_u v &= \frac{\langle u, v \rangle}{\|u\|^2} u = \frac{\langle (1, 2, 1), (-2, 1, 3) \rangle}{\|(-2, 1, 3)\|^2} (-2, 1, 3) \\ &= \frac{1(-2) + 2(1) + 1(3)}{14} (-2, 1, 3) \\ &= \frac{5}{14} (-2, 1, 3) \end{aligned}$$

We are now ready to find *orthogonal sets* of vectors and orthogonal bases. First we make an important definition.

Definition 2.4.4. Let V be a vector space with an inner product. A set of vectors $S = \{x_1, \dots, x_n\}$ in V is said to be *orthogonal* if $\langle x_i, x_j \rangle = 0$ for $i \neq j$. It is called **orthonormal** if also $\langle x_i, x_i \rangle = 1$. If, in addition, S is a basis it is called an orthogonal basis or orthonormal basis.

Note: Sometimes the conditions for orthonormality are written as

$$\langle x_i, x_j \rangle = \delta_{ij}$$

where δ_{ij} is the “Dirac” delta: $\delta_{ij} = 0$, $i \neq j$, $\delta_{ii} = 1$.

Theorem 2.4.1. Suppose U is a subspace of the (inner product) vector space V and that U has the basis $S = \{x_1 \dots x_k\}$, then U has an orthogonal basis.

Proof. Define $y_1 = \frac{x_1}{\|x_1\|}$. Thus y_1 is the “normalized” x_1 . Now define the new orthonormal basis recursively by

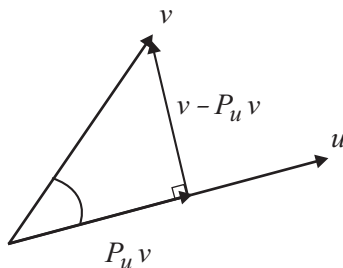
$$\begin{aligned} y'_{j+1} &= x_{j+1} - \sum_{i=1}^j \langle y_i, x_{j+1} \rangle y_i \\ y_{j+1} &= \frac{y'_{j+1}}{\|y'_{j+1}\|} \end{aligned}$$

for $j = 1, 2, \dots, k-1$. Then

- (1) y_{j+1} is orthogonal to $y_1, \dots, y_j$
- (2) $y_{j+1} \neq 0$.

In the language above we have $\langle y_i, y_j \rangle = \delta_{ij}$. □

Basically, what the proof accomplishes is to take the differences of the vector from the projections to the others. Referring to the figure above we compute $v - P_u v$ as noted in the figure below. The process of orthogonalization described above is called the **Gram–Schmidt** process.



Representation of vectors

One of the great advantages of orthonormal bases is that they make the representation of vectors particularly easy. It is as simple as computing an inner product. Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$ and with subspace U having basis $S = \{u_1, u_2, \dots, u_k\}$. Then for every $u \in U$ we know there are constants $a_1, a_2, \dots, a_k$ such that

$$x = a_1 u_1 + a_2 u_2 + \dots + a_k u_k.$$

Taking the inner product of both sides with u_j and applying the orthogonality relations

$$\begin{aligned} \langle x, u_j \rangle &= \langle a_1 u_1 + a_2 u_2 + \dots + a_k u_k, u_j \rangle \\ &= \sum_{i=1}^k a_i \langle u_i, u_j \rangle = a_j \end{aligned}$$

Thus $a_j = \langle x, u_j \rangle$, $j = 1, 2, \dots, k$, and

$$x = \sum_{j=1}^k \langle x, u_j \rangle u_j$$

Example 2.4.4. One basis of R_2 is given by the orthonormal vectors $S = \{u_1, u_2\}$, where $u_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]^T$ and $u_2 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]^T$. The representation of $x = [3, 2]^T$ is given by

$$x = \sum_{j=1}^2 \langle x, u_j \rangle u_j = \frac{5}{2} \sqrt{2} \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]^T + \frac{1}{2} \sqrt{2} \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]^T$$

Orthogonal subspaces

Definition 2.4.5. For any set of vectors S we define

$$S^\perp = \{v \in V \mid v \perp S\}$$

That is, $S^\perp$ is the set of vectors orthogonal to S . Often, $S^\perp$ is called the **orthogonal complement** or **orthocomplement** of S .

For example the orthocomplement of any vector $v = [v_1, v_2, v_3]^T \in R_3$ is the (unique) plane passing through the origin that is orthogonal to v . It is easy to see that the equation of the plane is $x_1v_1 + x_2v_2 + x_3v_3 = 0$.

For any set of vectors S the orthocomplement $S^\perp$ has the remarkable property of being a subspace of V , and therefore it must have an orthogonal basis.

Proposition 2.4.1. *Suppose that V is a vector space with an inner product, and $S \subset V$. Then $S^\perp$ is a subspace of V .*

Proof. If $y_1, \dots, y_m \in S^\perp$ then $\sum a_i y_i \in S^\perp$ for every set of coefficients $a_1, \dots, a_m$ in R (or $\mathbb{C}$). $\square$

Corollary 2.4.1. *Suppose that V is a vector space with an inner product, and $S \subset V$.*

(i) *If S is a basis of V , $S^\perp = \{0\}$.*

(ii) *If $U = \mathfrak{S}(S)$, then $U^\perp = S^\perp$.*

The proofs of these facts are elementary consequences of the proposition. An important decomposition result is based on orthogonality of subspaces. For example, suppose that V is a finite dimensional inner product space and that U is a subspace of V . Let $U^\perp$ be the orthocomplement of U , and let $S = \{u_1, u_2, \dots, u_k\}$ be an orthonormal basis of U . Let $x \in V$. Define $x_1 = \sum_{j=1}^k \langle x, u_j \rangle u_j$, and $x_2 = x - x_1$. Then it follows that $x_1 \in U$ and $x_2 \in U^\perp$. Moreover, $x = x_1 + x_2$. We summarize this in the following.

Proposition 2.4.2. *Let V is a vector space with inner product $\langle \cdot, \cdot \rangle$ and with subspace U . Then every vector $x \in V$ can be written as a sum of two orthogonal vectors $x = x_1 + x_2$, where $x_1 \in U$ and $x_2 \in U^\perp$.*

Geometrically what this results asserts is that for a given subspace of an inner product space, every vector has an orthogonal decomposition as two unique sum of a vector from the subspace and its orthocomplement. We write the vector components as the respective projections of the given vector to the orthogonal subspaces

$$\begin{aligned}x_1 &= P_U x \\x_2 &= P_{U^\perp} x\end{aligned}$$

Such decompositions are important in the analysis of vector spaces and matrices. In the case of vector spaces, of course, the representation of vectors is of great value. In the case of matrices, this type of decomposition serves to allow reductions of the matrices while preserving the information they carry.

2.4.1 An important equality for matrix multiplication and the inner product

Let $A \in M_{mn}(\mathbb{C})$. Then we know that both A^*A and AA^* (Alternatively, $A^T A$ and AA^T exist) exist, and we can surely inquire about the rank of these matrices. The main result of this section is on the rank of $A^T A$, namely that $r(A) = r(A^*A) = r(AA^*)$. The proof is quite simple but requires an important equality. Let $A \in M_{mn}(\mathbb{C})$ and $v \in \mathbb{C}_n$ and $w \in \mathbb{C}_m$. Then

$$\begin{aligned}\langle Av, w \rangle &= \left\langle \sum_{i=1}^m (Av)_i, w_i \right\rangle \\&= \sum_{i=1}^m \sum_{j=1}^n a_{ij} v_j \bar{w}_i \\&= \sum_{j=1}^n v_j \sum_{i=1}^m a_{ij} \bar{w}_i \\&= \sum_{j=1}^n v_j \overline{\sum_{i=1}^m \bar{a}_{ij} w_i} \\&= \sum_{j=1}^n v_j \overline{(A^* w)_j} \\&= \langle v, A^* w \rangle\end{aligned}$$

As a consequence we have $\langle A^*Av, w \rangle = \langle Av, Aw \rangle$ if both $v, w \in \mathbb{C}_n$. This important equality allows the adjoint or transpose matrices to be used on either side of the inner product, as needed. Indeed we shall use this below.

Proposition 2.4.3. *Let $A \in M_{mn}(\mathbb{C})$ have rank $r(A)$. Then*

$$r(A) = r(A^*A) = r(AA^*)$$

Proof. Assume that $r(A) = k$. Then there are k standard vectors $e_{j_1}, \dots, e_{j_k}$ such that for each $l = 1, 2, \dots, k$, the vectors Ae_{j_l} is one of the linearly independent columns of A . Moreover, it also follows that for every set of constants $a_1, \dots, a_k$ the vector $A(\sum a_l e_{j_l}) \neq 0$. Now $A^*A(\sum a_l e_{j_l}) \neq 0$ follows because

$$\begin{aligned} \left\langle A^*A \left(\sum a_l e_{j_l} \right), \left(\sum a_l e_{j_l} \right) \right\rangle &= \left\langle A \left(\sum a_l e_{j_l} \right), A \left(\sum a_l e_{j_l} \right) \right\rangle \\ &= \left\| A \left(\sum a_l e_{j_l} \right) \right\|^2 \neq 0 \end{aligned}$$

This in turn establishes that A^*A cannot be zero on a linear space of dimension k except for the zero element of course, and since the rank of A^*A cannot be larger than k the result is proved. $\square$

Remark 2.4.2. This establishes (6) of the Corollary 2.3.1 above. Also, it is easy to see that the result is also true for real matrices.

2.4.2 The Legendre Polynomials

When a vector space has an inner product, it is possible to construct an orthogonal basis from any given basis. We do this now for the polynomial space $P_n(-1, 1)$ and a particular basis.

Consider the space the polynomials of degree n defined on the interval $[-1, 1]$ over the reals. Recall that this is a vector space and has as a basis the *monomials* $\{1, x, x^2, \dots, x^n\}$. We can define an assortment of inner products on this space, but the most common inner product is given by

$$\langle p, q \rangle = \int_{-1}^1 p(x) q(x) \, dx$$

Verifying the inner product properties is fairly straight forward and we leave it as an exercise. This inner product also defines a norm

$$\|p\|^2 = \int_{-1}^1 |p(x)|^2 \, dx$$

This norm satisfies the triangle inequality requires an integral version of the Cauchy-Schwartz inequality.

Now that we have an inner product and norm, we could proceed to find an orthogonal basis of $P_n(-1, 1)$ by applying the Gram-Schmidt procedure to the basis $\{1, x, x^2, \dots, x^n\}$. This procedure can be clumsy and tedious. It is easier to build an orthogonal basis from scratch. Following tradition we will use capital letters $P_0, P_1, \dots$ to denote our orthogonal polynomials. Toward this end take $P_0 = 1$. Note we are numbering from 0 onwards so that the polynomial degree will agree with the index. Now let $P_1 = ax + b$. For orthogonality, we need

$$\int_{-1}^1 P_0(x) P_1(x) dx = \int_{-1}^1 1 \cdot (ax + b) dx = 2b = 0$$

Thus $b = 0$ and a can be arbitrary. We take $a = 1$. This gives $y_1 = x$. Now we assume the model for the next orthogonal function to be $y_2 = ax^2 + bx + c$. This time there are two orthogonality conditions to satisfy.

$$\begin{aligned} \int_{-1}^1 P_0(x) P_2(x) dx &= \int_{-1}^1 1 \cdot (ax^2 + bx + c) dx = \frac{2}{3}a + 2c = 0 \\ \int_{-1}^1 P_1(x) P_2(x) dx &= \int_{-1}^1 x \cdot (ax^2 + bx + c) dx = \frac{2}{3}b = 0 \end{aligned}$$

We conclude that $b = 0$. From the equation $\frac{2}{3}a + 2c = 0$, we can assign one of the variables and solve for the other one. Following tradition we take $c = -\frac{1}{2}$ and solve for a to get $a = \frac{3}{2}$.

The next polynomial will be modeled as $P_3(x) = ax^3 + bx^2 + cx + d$. Three orthogonality relations need to be satisfied.

$$\begin{aligned} \int_{-1}^1 P_0(x) P_3(x) dx &= \int_{-1}^1 1 \cdot (ax^3 + bx^2 + cx + d) dx = \frac{2}{3}b + 2d = 0 \\ \int_{-1}^1 P_1(x) P_3(x) dx &= \int_{-1}^1 x \cdot (ax^3 + bx^2 + cx + d) dx = \frac{2}{5}a + \frac{2}{3}c = 0 \\ \int_{-1}^1 P_2(x) P_3(x) dx &= \int_{-1}^1 \frac{1}{2}(3x - 1)(ax^3 + bx^2 + cx + d) dx \\ &= \frac{3}{5}a - \frac{1}{3}b + c - d = 0 \end{aligned}$$

It is easy to see that $b = d = 0$ (why?) and from $\frac{2}{5}a + \frac{2}{3}c = 0$, we select

$c = -\frac{3}{2}$ and $a = \frac{5}{2}$. Our table of orthogonal polynomials so far is

k	$P_k(x)$
0	1
1	x
2	$\frac{1}{2}(3x - 1)$
3	$\frac{1}{2}(5x^3 - 3x)$

Continue in this fashion, generating polynomials of increasing order each orthogonal to all of the lower order ones.

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \frac{3}{2}x^2 - \frac{1}{2} \\
P_3(x) &= \frac{5}{2}x^3 - \frac{3}{2}x \\
P_4(x) &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8} \\
P_5(x) &= \frac{63}{8}x^5 - \frac{35}{4}x^3 + \frac{15}{8}x \\
P_6(x) &= \frac{231}{16}x^6 - \frac{315}{16}x^4 + \frac{105}{16}x^2 - \frac{5}{16} \\
P_7(x) &= \frac{429}{16}x^7 - \frac{693}{16}x^5 + \frac{315}{16}x^3 - \frac{35}{16}x \\
P_8(x) &= \frac{6435}{128}x^8 - \frac{3003}{32}x^6 + \frac{3465}{64}x^4 - \frac{315}{32}x^2 + \frac{35}{128} \\
P_9(x) &= \frac{12155}{128}x^9 - \frac{6435}{32}x^7 + \frac{9009}{64}x^5 - \frac{1155}{32}x^3 + \frac{315}{128}x \\
P_{10}(x) &= \frac{46189}{256}x^{10} - \frac{109395}{256}x^8 + \frac{45045}{128}x^6 - \frac{15015}{128}x^4 + \frac{3465}{256}x^2 - \frac{63}{256}
\end{aligned}$$

2.4.3 Orthogonal matrices

Besides sets of vectors being orthogonal, there is also a definition of orthogonal matrices. The two notions are closely linked.

Definition 2.4.6. We say a matrix $A \in M_n(\mathbb{C})$ is *orthogonal* if $A^*A = I$. The same definition applies to matrices $A \in M_n(R)$ with A^* replaced by A^T .

For example, the rotation matrices (Exercise ??) $B_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ are all orthogonal.

A simple consequence of this definition is that the rows and the columns of A are **orthonormal**. We see for example that when A is orthogonal then $(A^*)^2 = (A^{-1})^2 = A^{-1}A^{-1} = (A^2)^{-1}$. Such a definition applies, as well to higher powers. For instance, if A is orthogonal then A^m is orthogonal for every positive integer m .

One way to generate orthogonal matrices in $\mathbb{C}_n$ (or R_n) is to begin with an orthonormal basis and arrange it into an $n \times n$ matrix either as its columns or rows.

Theorem 2.4.2. (i) Let $\{x_i\}$, $i = 1, \dots, n$ be an orthonormal basis of $\mathbb{C}_n$ or (R_n) . Then the matrices

$$U = \begin{bmatrix} x_1 & \cdots & x_n \\ \downarrow & \cdots & \downarrow \\ \cdot & & \cdot \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} x_1 & \longrightarrow & \cdot \\ \vdots & \vdots & \\ x_n & \longrightarrow & \cdot \end{bmatrix}$$

formed by arranging the vectors x_i as its respective columns or rows are orthogonal.

(ii) Conversely, U is an orthogonal matrix, the sets of its rows and columns are each orthonormal, and moreover each forms a basis of $\mathbb{C}_n$ or (R_n) .

The proofs are entirely trivial. We shall consider these types of results in more detail later in Chapter 4. In the meantime there are a few more interesting results that are direct consequences of the definition and facts about the transpose (adjoint).

Theorem 2.4.3. Let $A, B \in M_n(C)$ (or $M_n(R)$) be orthogonal matrices. Then

- (a) A is invertible and $A^{-1} = A^*$.
- (b) For each integer $k = 0, \pm 1, \pm 2, \dots$, both A^k and $-A^k$ are orthogonal.
- (c) AB is orthogonal.

2.5 Determinants

This section is about determinants that can be regarded as a measure of singularity of a matrix. More generally, in many applied situations that deal with complex objects, a single number is sought that will in some way

classify an aspect of those objects. The determinant is such a measure for singularity of the matrix. The determinant is difficult to calculate and of not much practical use. However, it has considerable theoretical value and certainly has a place of historical interest.

Definition 2.5.1. Let $A \in M_n(F)$. Define the **determinant** of A to be the value in F

$$\det A = \sum_{\sigma} \left(\prod_{i=1}^n a_{i\sigma(i)} \right) \cdot \operatorname{sgn} \sigma$$

where σ is a permutation of the integers $\{1, 2, \dots, n\}$ and

(1) $\sum_{\sigma}$ denotes the sum over all permutations

(2) $\operatorname{sgn} \sigma = \text{sign of } \sigma = \pm 1$

A **transposition** is the exchange of two elements of an ordered list with all others staying the same. With respect to permutations, a transposition of one permutation is another permutation formed by the exchange of two values. For example a transposition of $\{1, 4, 3, 2\}$ is $\{1, 3, 4, 2\}$. The **sign** of a given permutation σ is

(a) $+1$, if the number of transpositions required to bring σ to $\{1, 2, \dots, n\}$ is even.

(b) -1 , if the number of transpositions required to bring σ to $\{1, 2, \dots, n\}$ is odd.

Alternatively, and what is the same thing, we may count the number m of transpositions required to bring σ to $\{1, 2, \dots, n\}$ and to compute the sign is $(-1)^m$.

Example 2.5.1.

$$\begin{aligned} \sigma_1 = \{2, 1, 3\} & \xrightarrow{1 \leftrightarrow 2} \{1, 2, 3\} & \text{odd} \\ \sigma_2 = \{2, 3, 1\} & \xrightarrow{3 \leftrightarrow 1} \{2, 1, 3\} \xrightarrow{1 \leftrightarrow 2} \{1, 2, 3\} & \text{even} \\ \operatorname{sgn} \sigma_1 &= -1 & \operatorname{sgn} \sigma_2 = +1 \end{aligned}$$

Proposition 2.5.1. Let $A \in M_n$.

(i) If two rows of A are interchanged to obtain B , then

$$\det B = -\det A.$$

(ii) Given $A \in M_n(F)$. If any row is multiplied by a scalar c , the resulting matrix B has determinant

$$\det B = c \det A.$$

(iii) If any two rows of $A \in M_n(F)$ are equal,

$$\det A = 0.$$

Proof. (i) Suppose rows i_1 and i_2 are interchanged. Now for the given permutations σ apply the transposition $i_1 \leftrightarrow i_2$ to get σ_1 . Then

$$\prod_{i=1}^n a_{i_1\sigma(i)} = \prod_{i=1}^n b_{i_2\sigma_1(i)}$$

because

$$a_{i_1\sigma(i_1)} = b_{i_2\sigma_1(i_2)}$$

as

$$b_{i_2j} = a_{i_1j} \quad \text{and} \quad \sigma_1(i_2) = \sigma_2(i_1)$$

and similarly $a_{i_2\sigma(i_2)} = b_{i_1\sigma_1(i_1)}$. All other terms are equal. In the computation of the full determinant with signs of the permutations, we see that the change is caused only by the fact $\text{sgn}(\sigma_1) = -\text{sgn}(\sigma)$. Thus,

$$\det B = -\det A.$$

(ii) Is trivial.

(iii) If two rows are equal then by part (i)

$$\det A = -\det A$$

and this implies $\det A = 0$.

□

Corollary 2.5.1. *Let $A \in M_n$. If A has two rows equal up to a multiplicative constant, it has determinant zero.*

What happens to the determinant when two matrices are added. The result is too complicated to write down is not very important. However, when a single vector is added to a row or a column of a matrix, then the result can be simply stated.

Proposition 2.5.2. *Suppose $A \in M_n(F)$. Suppose B is obtained from A by adding a vector v to a given row (resp. column) and C is obtained from A by replacing the given row (resp. column) by the vector v . Then*

$$\det B = \det A + \det C.$$

Proof. Assume the j^{th} row is altered. Using the definition of the determinant,

$$\begin{aligned} \det A &= \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_i b_{i\sigma(i)} = \sum_{\sigma} \operatorname{sgn}(\sigma) \left(\prod_{i \neq j} b_{i\sigma(i)} \right) b_{j\sigma(j)} \\ &= \sum_{\sigma} \operatorname{sgn}(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) (a + v)_{j\sigma(j)} \\ &= \sum_{\sigma} \operatorname{sgn}(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) a_{j\sigma(j)} + \sum_{\sigma} \operatorname{sgn}(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) v_{j\sigma(j)} \\ &\quad \det A + \det C \end{aligned}$$

For column replacement the proof is similar, particularly using the alternate representation of the determinant given in Exercise 15. $\square$

Corollary 2.5.2. *Suppose $A \in M_n(F)$ and B is obtained by multiplying a given row (resp. column) of A by a scalar and adding it to another row (resp. column), then*

$$\det B = \det A.$$

Proof. First note that in applying Proposition 2.5.2 C has two rows equal up to a multiplicative constant. Thus $\det C = 0$. $\square$

Computing determinants is usually difficult and many techniques have been devised to compute them out. As is evident from counting, computing

the determinant of an $n \times n$ matrix using the definition above would require the expression of all $n!$ permutations of the integers $\{1, 2, \dots, n\}$ and the determination of their signs together with all the concomitant products and summation. This method is prohibitively costly. Using elementary row operations and Gaussian elimination, the evaluation of the determinant becomes more manageable. First we need the result below.

Theorem 2.5.1. *For the elementary matrices the following results hold.*

(a) for Type 1 (row interchange) E_1

$$\det E_1 = -1$$

(b) for Type 2 (multiply a row by a constant c) E_2

$$\det E_2 = c$$

(c) for Type 3 (add a multiple of one row to another row) E_3

$$\det E_3 = 1.$$

Note that (c) is a consequence of Corollary 2.5.2. Proof of parts (a) and (b) are left as exercises. Thus for any matrix $A \in M_n(F)$ we have

$$\begin{aligned} \det(E_1 A) &= -\det A &= \det E_1 \det A \\ \det(E_2 A) &= c \det A &= \det E_2 \det A \\ \det(E_3 A) &= \det A &= \det E_3 \det A. \end{aligned}$$

Suppose $F_1 \dots F_k$ is a sequence of row operations to reduce A to its RREF. Then

$$F_k F_{k-1} \dots F_1 A = B.$$

Now we see that

$$\begin{aligned} \det B &= \det(F_k F_{k-1} \dots F_1 A) \\ &= \det(F_k) \det(F_{k-1} \dots F_1 A) \\ &= \quad \vdots \\ &= \det(F_k) \det(F_{k-1}) \dots \det(F_1) \det A. \end{aligned}$$

For B in RREF and $B \in M_n(F)$, we have that B is upper triangular. The next result establishes Theorem 2.3.4(k) about the determinant of singular and non singular matrices. Moreover, the determinant of triangular matrices is computed simply as the product of its diagonal elements.

Proposition 2.5.3. *Let $A \in M_n$. Then*

(i) *If $r(A) < n$, then $\det A = 0$.*

(ii) *If A is triangular then*

$$\det A = \prod a_{ii}$$

(iii) *If $r(A) = n$, then $\det A \neq 0$.*

Proof. (i) If $r(A) < n$, then its RREF has a row of zeros, and $\det A = 0$ by Theorem 2.5.1. (ii) If A is triangular the only product without possible zero entries is $\prod a_{ii}$. Hence $\det A = \prod a_{ii}$. (iii) If $r(A) = n$, then its RREF has no nonzero rows. Since it is square and has a leading one in each column, it follows that the RREF is the identity matrix. Therefore $\det A \neq 0$. $\square$

Now let $A, B \in M_n$. If A is singular the RREF must have a zero row. It follows that $\det A = 0$. If A is singular it follows that AB is singular. Therefore

$$0 = \det AB = \det A \det B.$$

The same reasoning applies if B is singular. If A and B are not singular both A and B can be row reduced to the identity. Let $F_1 \dots F_{k_A}$ be the row operations that reduce A to I , and $G_1 \dots G_{k_B}$ be the row operations that reduce B to I . Then

$$\begin{aligned} \det A &= [\det(F_1) \dots \det(F_{k_A})]^{-1} \\ \det B &= [\det(G_1) \dots \det(G_{k_B})]^{-1}. \end{aligned}$$

Also

$$I = (G_{k_B} \dots G_1)(F_{k_B} \dots F_1)AB$$

and we have

$$\det I = (\det A)^{-1}(\det B)^{-1} \det AB.$$

This proves the

Theorem 2.5.2. *If $A, B \in M_n(F)$, $\det AB = \det A \det B$.*

2.5.1 Minors and Determinants

The method of row reduction is one of the simplest methods to compute the determinant of a matrix. Indeed, it is not necessary to use Type 2 elementary transformation. This results in the computing the determinant as the product of the diagonal elements of the resulting triangular matrix possibly multiplied by a minus sign. An alternate approach to computing determinants using *minors* is both interesting and useful. However, unless the matrix has some special form, it does not provide a computational alternative to row reduction.

Definition 2.5.2. Let $A \in M_n(C)$. For any row i and column j define the (ij) -minor of A by

$$M_{ij} = \det A \left| \begin{array}{c} i^{\text{th}} \text{ row removed} \\ j^{\text{th}} \text{ column removed} \end{array} \right|$$

The notation

$$A \left| \begin{array}{c} i^{\text{th}} \text{ row removed} \\ j^{\text{th}} \text{ column removed} \end{array} \right|$$

denotes the $(n-1) \times (n-1)$ matrix formed from A by removing the i^{th} row and j^{th} column. With minors an alternative formulation of the determinant can be given. This method, while not of great value computationally, has some theoretical importance. For example, the inverse of a matrix can be expressed using minors. We begin by consideration the determinant.

Theorem 2.5.3. Let $A \in M_n(C)$. (i) Fix any row, say row k . The determinant of A is given by

$$\det A = \sum_{j=1}^n a_{kj} (-1)^{k+j} M_{kj}$$

(ii) Fix any column, say column m . The determinant of A is given by

$$\det A = \sum_{j=1}^n a_{jm} (-1)^{m+j} M_{jm}$$

Proof. (i) Suppose that $k = 1$. Consider the quantity

$$a_{11}M_{11}$$

We observe that this is equivalent to all the products of the form

$$\operatorname{sgn}(\sigma) a_{11} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

where only permutations that fix the integer (i.e. position) 1 are taken. Thus $\sigma(1) = 1$. Since this position is fixed the signs taken in the determinant M_{11} for permutations of $n - 1$ integers are respectively the same as the signs for the new permutation of n integers.

Now consider all permutations that fix the integer 2 in the sense that $\sigma(1) = 2$. The quantity $a_{12}(-1)^{1+2} M_{12}$ consists of all the products of the form

$$\operatorname{sgn}(\sigma) a_{12} \cdot a_{2\sigma(1)} a_{3\sigma(3)} \cdots a_{n\sigma(n)}$$

We need here the extra sign change because if the part of the permutation σ of the integers $\{1, 3, 4, \dots, n\}$ is of one sign, which is the sign used in the computation of $\det M_{ij}$, then the permutation of σ of the integers $\{1, 2, 3, 4, \dots, n\}$ is of the other sign, and that sign is $\operatorname{sgn}(\sigma)$.

When we proceed to the k^{th} component, we consider permutations that fix the integer k . That is, $\sigma(1) = k$. In this case the quantity $a_{1k}(-1)^{1+k} M_{1k}$ consists of all products of the form

$$\operatorname{sgn}(\sigma) a_{1k} a_{2\sigma(1)} \cdots a_{k-1\sigma(k-1)} a_{k+1\sigma(k+1)} \cdots a_{n\sigma(n)}$$

Continuing in this way we exhaust all possible products $a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ over all possible permutations of the integers $\{1, 2, \dots, n\}$. This proves the assertion. The proof for expanding from any row is similar, with only a possible change of sign needed, which is a prescribed.

(ii) The proof is similar. □

Example 2.5.2. Find the determinant of

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 0 & 1 & 3 \\ 1 & 2 & -1 \end{bmatrix}$$

expanding across the first row and then expanding down the second column.

Solution. Expanding across the first row gives

$$\begin{aligned} \det A &= a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} \\ &= 3 \det \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} - 2 \det \begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix} - 1 \det \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \\ &= 3(-7) - 2(-3) - (-1) \\ &= -14 \end{aligned}$$

Expanding across the second column gives

$$\begin{aligned}\det A &= -2 \det \begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix} + 1 \det \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} - 2 \det \begin{bmatrix} 3 & -1 \\ 0 & 3 \end{bmatrix} \\ &= -2(-3) + (-2) - 2(9) = -14\end{aligned}$$

The inverse of the matrix can be formulated in terms of minors, which is formulated below.

Definition 2.5.3. Let $A \in M_n(\mathbb{C})$ (or $M_n(R)$). Define the **adjugate** (or **adjoint**) matrix $\hat{A}$ by

$$\hat{A}_{ij} = (-1)^{i+j} M_{ji}$$

where M_{ji} is the ji minor.

The adjugate has traditionally been called the “adjoint”, but that terminology is somewhat ambiguous in light of the previous definition as complex conjugate transpose. Note that it is defined for all square matrices; when restricted to invertible matrices the inverse appears.

Theorem 2.5.4. Let $A \in M_n(\mathbb{C})$ (or $M_n(R)$) be invertible. Then $A^{-1} = \frac{1}{\det A} \hat{A}$

Proof. A quick examination of the ij -entry of the product $A\hat{A}$ yields the following sum

$$\sum_{j=1}^n a_{ij} \hat{A}_{jk} = \frac{1}{\det(A)} \sum_{j=1}^n a_{ij} (-1)^{k+j} M_{kj}$$

There are two possibilities. (1) If $i = k$, then the summation above is the summation to form the determinant as described in Theorem 2.5.3. (2) If $i \neq k$, the summation is the computation of the determinant of the matrix A with the k^{th} row replaced by the i^{th} row. Thus the determinant of a matrix with two identical rows is represented above and this must be zero. We conclude that $(A\hat{A})_{ij} = \delta_{ij}$, the usual Kronecker ‘delta,’ and the result is proved. $\square$

Example 2.5.1. Find the adjugate and inverse of

$$A = \begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix}$$

It is easy to see that

$$\hat{A} = \begin{bmatrix} 1 & -4 \\ -2 & 2 \end{bmatrix}$$

Also $\det A = -6$. Therefore, the inverse

$$A^{-1} = -\frac{1}{6} \begin{bmatrix} 1 & -4 \\ -2 & 2 \end{bmatrix}$$

Remark 2.5.1. The notation for cofactors of a square matrix A is often used

$$\hat{a}_{ij} = (-1)^{i+j} M_{ji}$$

Note the reversed order of the subscripts ij and then ji above.

Cramer's Rule

We know now that the solution to the system $Ax = b$ is given by $x = A^{-1}b$.

Moreover, the inverse A^{-1} is given by $A^{-1} = \frac{\hat{A}}{\det A}$, where $\hat{A}$ is the adjugate matrix. The i^{th} component of the solution vector is therefore

$$\begin{aligned} x_i &= \frac{1}{\det A} \sum_{j=1}^n \hat{a}_{ij} b_j \\ &= \frac{1}{\det A} \sum_{j=1}^n (-1)^{i+j} M_{ji} b_j \\ &= \frac{\det A_i}{\det A} \end{aligned}$$

where we define the matrix A_i to be the modification to A by replacing its i^{th} column by the vector b . In this way we obtain a very compact formula for the solution of a linear system. Called Cramer's rule we state this conclusion as

Theorem 2.5.1. (*Cramer's Rule.*) Let $A \in M_n(\mathbb{C})$ be invertible and $b \in \mathbb{C}^n$. For each $i = 1, \dots, n$, define the matrix A_i to be the modification of A by replacing its i^{th} column by the vector b . Then the solution to the linear system $Ax = b$ is given by components $x_i = \frac{\det A_i}{\det A}$, $i = 1, \dots, n$.

Example 2.5.2. Given the matrix $A = \begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix}$, and the vector $b = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Solve the system $Ax = b$ by Cramer's rule.

We have

$$A_1 = \begin{bmatrix} 2 & 4 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

and $\det A_1 = 6$, $\det A_2 = -6$, $\det A = -6$. Therefore

$$x_1 = -1 \quad \text{and} \quad x_2 = 1$$

A curious formula

The useful formula using cofactors given below will have some consequence when we study positive definite operators in Chapter ??.

Proposition 2.5.1. *Consider the matrix*

$$B = \begin{bmatrix} 0 & x_1 & x_2 & \cdots & x_n \\ x_1 & a_{11} & a_{12} & \cdots & a_{1n} \\ x_2 & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}$$

Then

$$\det B = - \sum \hat{a}_{ij} x_i x_j$$

where $\hat{a}_{ij}$ is the ij -cofactor of A .

Proof. Expand by minors along the top row to get

$$\det B = \sum (-1)^j x_j M_{1j}(B)$$

Now expand the matrix of $M_{1j}(B)$ down the first column. This gives

$$M_{1j}(B) = \sum (-1)^{i-1} x_i M_{ij}(A)$$

Combining we obtain

$$\begin{aligned}
 \det B &= \sum (-1)^j x_j M_{1j}(B) \\
 &= \sum (-1)^j x_j \left[\sum (-1)^{i-1} x_i M_{ij}(A) \right] \\
 &= \sum \sum (-1)^{i+j-1} x_j x_i M_{ij}(A) \\
 &= - \sum \sum \hat{a}_{ij} x_j x_i
 \end{aligned}$$

□

The reader may note that in the last line of the equation above, we should have used $\hat{a}_{ij}$. However, the formulation given is correct, as well. (Why?)

2.6 Partitioned Matrices

It is convenient to study partitioned or “blocked” matrices, or more graphically said, matrices whose entries are themselves matrices. For example, with I_2 denoting the 2×2 identity matrix we can create the 4×4 matrix written in partitioned form and expanded form.

$$A = \begin{bmatrix} aI_2 & cI_2 \\ cI_2 & dI_2 \end{bmatrix} = \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix}$$

Partitioning matrices allows our attention to focus on certain structural properties. In many applications partitioned matrices appear in a natural way, with the particular blocks having some system context. Many similar subclasses and processes apply to partitioned matrices. In specific situations they can be added, multiplied, and inverted, just like regular matrices. It is even possible to perform “blocked” version of Gaussian elimination. In the few results here, we touch on some of these possibilities.

Definition 2.6.1. For each $1 \leq i \leq m$ and $1 \leq j \leq n$, let A_{ij} be an $m_i \times n_j$ matrices where . Then the matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{2n} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

is a partitioned matrix of order $(\sum m)_i \times (\sum n_j)$.

The usual operations of addition and multiplication of partitioned matrices can be performed provided each of the operations makes sense. For addition of two partitioned matrices A and B it is necessary to have the same numbers of blocks of the respective same sizes. Then

$$\begin{aligned} A + B &= \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{2n} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{2n} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{bmatrix} \\ &= \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1n} + B_{1n} \\ A_{21} + B_{21} & A_{2n} + B_{2n} & \cdots & A_{2n} + B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} + B_{m1} & A_{m2} + B_{m2} & \cdots & A_{mn} + B_{mn} \end{bmatrix} \end{aligned}$$

For multiplication, the situation is a bit more complicated. For definiteness, suppose that B is a partitioned matrix with block sizes $s_i \times t_j$, where $1 \leq i \leq p$ and $1 \leq j \leq q$. The usual operations to construct $C = AB$,

$$\sum_{j=1}^n A_{ij} B_{jk}$$

then make sense provided $p = n$ and $n_j = s_j$, $1 \leq j \leq n$.

A special category of partitioned matrices are the so-called **quasi-triangular** matrices, wherein $A_{ij} = 0$ if $i > j$ for the “lower” triangular version. The special subclass of quasi-triangular matrices wherein $A_{ij} = 0$ if $i \neq j$ are called **quasi-diagonal**. In the case of the multiplication of partitioned matrices ($C = AB$) with the left multiplicand A a quasi-diagonal matrix, we have $C_{ik} = A_{ii} B_{ik}$. Thus the multiplication is similar in form to the usual multiplication of matrices where the left multiplicand is a diagonal matrix. In the case of the multiplication of partitioned matrices ($C = AB$) with the right multiplicand B a quasi-diagonal matrix, we have $C_{ik} = A_{ik} B_{kk}$. For quasi-triangular matrices with square diagonal blocks, there is an interesting result about the determinant.

Theorem 2.6.1. *Let A be a quasi-triangular matrix, where the diagonal blocks A_{ii} are square. Then*

$$\det A = \prod_i \det A_{ii}$$

Proof. Apply row operations on each vertical block without row interchanges between blocks, without any Type 2 operations. The resulting matrix in each diagonal block position (i, i) is triangular. Be sure to multiply one of the diagonal entries by ± 1 , reflecting the number of row interchanges within a block. The resulting matrix can still be regarded as partitioned, though the diagonal blocks are now actually upper triangular. Now apply Proposition 2.5.3, noting that the product of each of the diagonal entries pertaining to the i^{th} block is in fact $\det A_{ii}$. $\square$

A simple consequence of this result, proved alá Gaussian elimination, is contained in the following corollary.

Corollary 2.6.1. *Consider the partitioned matrix*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with square diagonal blocks and with A_{11} invertible. Then the rank of A is the same as the rank of A_{11} if and only if $A_{22} = A_{21}A_{11}^{-1}A_{12}$.

Proof. Multiplication of A by the elementary partitioned matrix

$$E = \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}$$

yields

$$\begin{aligned} EA &= \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \end{aligned}$$

Since E has full rank, it follows that $\text{rank}(EA) = \text{rank} A$. Since EA is quasi-triangular, it follows that the rank of A is the same as the rank of A_{11} if and only if $A_{22} - A_{21}A_{11}^{-1}A_{12} = 0$. $\square$

2.7 Linear Transformations

Definition 2.7.1. A mapping T from R_n to R_m is called a **linear transformation** if

$$\begin{aligned} T(x + y) &= Tx + Ty & \forall x, y \in R_n \\ T(ax) &= aTx & \forall a \in F. \end{aligned}$$

Note: We normally write Tx instead of $T(x)$.

Example 2.7.1. $T: R_n \rightarrow R_m$. Let $a \in R_m$ and $y \in R_n$. Then for each $x \in R_n$, $Tx = \langle x, y \rangle a$ is a linear transformation. Let $S = \{v_1 \dots v_n\}$ be a basis of R_n , and define the $m \times n$ matrix with columns given by the coordinates of $Tv_1, Tv_2, \dots, Tv_n$. Then this matrix

$$A = \begin{bmatrix} Tv_1 & Tv_2 & \dots & Tv_n \\ \downarrow & \downarrow & \ddots & \downarrow \end{bmatrix}$$

is the matrix representation of T with respect to the basis S . Thus, if $x = \sum a_i v_i$, whence $[x]_S = (a_1 \dots a_n)$, we have

$$[Tx]_S = A[x]_S$$

★ There is a duality between all linear transformations from R_n to R_m and the set $M_{m,n}(F)$.

Note that $M_{m,n}(F)$ is itself a vector space over F . Hence $\mathcal{L}(F_n, F_m)$, the set of linear transformations from F_n to F_m is likewise. As such it has subspaces.

Example 2.7.2. (1) Let $\bar{x} \in F_n$. Define $\mathcal{J} = \{T \in \mathcal{L} \mid T\bar{x} = 0\}$. Then $\mathcal{J}$ is a subspace of $\mathcal{L}(F_n, F_m)$.

(2) Let $\mathcal{U} = \{T \in \mathcal{L}(R_n, R_n) \mid Tx \geq 0 \text{ if } x \geq 0\}$, where $\{x \geq 0\}$ means the positive orthant of R_n . $\mathcal{U}$ is not a linear subspace of $\mathcal{L}(R_n, R_n)$, though it is a convex set.

(3) Define $T: P_n \rightarrow P_n$ by $Tp = \frac{d}{dx}p$. T is a linear transformation.

Example 2.7.3. Express the linear transformation $D: P_3 \rightarrow P_3$ given by $Dp = \frac{d}{dx}p(x)$ as a matrix with respect to the basis. $S = \{1, x, x^2, x^3\}$. We have $D1 = 0 = 0 + 0x + 0x^2 + 0x^3$. Also

$$[D1]_S = [0, 0, 0, 0]^T$$

similarly

$$\begin{aligned} [Dx]_S &= [1, 0, 0, 0]^T \\ [Dx^2]_S &= [0, 2, 0, 0]^T \\ [Dx^3]_S &= [0, 0, 3, 0]^T. \end{aligned}$$

Hence

$$[D]_S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In this context the differentiation operator is rather simple.

Example 2.7.4. Consider the linear transformation T defined by $Tq = 3x\frac{d}{dx}q + x^2q$ for $q \in P_2$. Find the matrix representation of T .

Solution. First off we notice that this transformation has range in P_4 . Let's use the standard bases for this problem. We then determine the coordinates of T for vectors in the P_2 basis $\{1, x, x^2\}$ in the P_4 basis $\{1, x, x^2, x^3, x^4\}$. Compute

$$\begin{aligned} T(1) &= x^2 \\ T(x) &= 3x + x^3 \\ T(x^2) &= 6x^2 + x^4 \end{aligned}$$

The coordinates of the input vectors we know are $[1, 0, 0]^T$, $[0, 1, 0]^T$, and $[0, 0, 1]^T$. For the output vectors the coordinates are $[0, 0, 1, 0, 0]^T$, $[0, 3, 0, 1, 0]^T$, and $[0, 0, 6, 0, 1]^T$. So, with respect to these two bases, the matrix of the transformation is

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 0 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Observe that the dimensionality corresponds with the dimensionality of the respective spaces.

Example 2.7.5. Let $V = R_2$, with $S_0 = \{v_1, v_2\} = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$, $S_1 = \{w_1, w_2\} = \left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}\right\}$, and $T = I$ the identity. The vectors above are expressed in the standard $E = \{e_1, e_2\}$, $Tv_j = Iv_j = v_j$. To find $[v_j]_{S_1}$ we solve

$$v_j = a_j w_1 + \beta_j w_2$$

$$\begin{aligned}
v_1: \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \longrightarrow \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ -\frac{2}{5} \end{bmatrix} \\
v_2: \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \longrightarrow \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ -\frac{1}{5} \end{bmatrix}
\end{aligned}
\quad \begin{array}{l} \nwarrow \text{solve linear} \\ \swarrow \text{systems} \end{array}$$

Therefore

$$s_1[I]_{S_0} = \begin{bmatrix} \frac{1}{5} & \frac{3}{5} \\ -\frac{2}{5} & -\frac{1}{5} \end{bmatrix} \quad \leftarrow \begin{array}{l} \text{change of} \\ \text{basis} \\ \text{matrix} \end{array}$$

If

$$\begin{aligned}
[x]_{S_0} &= \begin{bmatrix} -1 \\ 2 \end{bmatrix} & [x]_{S_1} &= s_1[I]_{S_0} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\
& & &= \frac{1}{5} \begin{bmatrix} 1 & 3 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.
\end{aligned}$$

Note the necessity of using the standard basis to express the vectors in both bases S_0 and S_1 .

2.8 Change of Basis

Let V be a vector space with bases $S_0 = \{v_1 \dots v_n\}$ and $S_1 = \{w_1 \dots w_n\}$, and suppose $T: V \rightarrow V$ is a linear transformation. We want to find the representation of T as a matrix that takes a vector x given in terms of its S_0 coordinates and produces the vector Tx given in terms of its S_1 coordinates.

We know that $x \rightarrow [x]_{S_0}$ is well defined. The action of T is known if the n vector $[x]_{S_0} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ and the vectors $Tv_1, Tv_2, \dots, Tv_n$ are known, for if $x = \sum c_j v_j$, then $Tx = \sum c_j Tv_j$, by linearity.

To determine $[Tx]_{S_1}$ we need to convert the Tv_j , $j = 1, \dots, n$ to coordinates in the other S_1 basis. This is done as follows. Find

$$[Tv_j]_{S_1} = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{bmatrix} \quad j = 1, 2, \dots, n.$$

Then if $x \in V$

$$\begin{aligned} [Tx]_{S_1} &= [\Sigma c_j T v_j]_{S_1} = \Sigma c_j [T v_j]_{S_1} \\ &= \left[\sum_j t_{ij} c_j \right] \\ &= \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ t_{n1} & & t_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}. \end{aligned}$$

This $n \times n$ array $[t_{ij}]$ depends on T, S_0 and S_1 but not on x . We define the $S_0 \rightarrow S_1$ basis representation of T to be $[t_{ij}]$, and we write this as

$$_{S_1}[T]_{S_0} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix}.$$

In the special case that T is the identity operator the matrix $_{S_1}[I]_{S_0}$ converts the coordinates of a vector in the basis S_0 to coordinates in the basis S_1 . It is easy to see that $_{S_0}[I]_{S_1}$ must be the inverse of $_{S_1}[I]_{S_0}$ and thus

$$_{S_0}[I]_{S_1} \cdot _{S_1}[I]_{S_0} = I.$$

We can also establish the equality

$$_{S_1}[T]_{S_1} = _{S_1}[I]_{S_0} _{S_0}[T]_{S_0} _{S_0}[I]_{S_1}.$$

In this way we see that the matrix representation of T depends on the bases involved. If X is any invertible matrix in $M_n(F)$ we can write

$$B = X^{-1}AX.$$

The interpretation in this context is clear

- X : change of coordinate from one basis to another $S_0 \rightarrow S_1$
- X^{-1} : change of coordinate $S_1 \rightarrow S_0$
- A : matrix of the linear transformation in the basis S_0
- B : matrix of the same linear transformation in the basis S_1 .

With this in mind it seems prudent to study linear transformations in the basis that makes their matrix representation as simple as possible.

Example 2.8.1. Let $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$ be the matrix representation of a linear transformation given with respect to the standard basis $S_0 = \{e_1, e_2\} = \{(1, 0), (0, 1)\}$. Find the matrix representation of this transformation with respect to the basis $S_1 = \{v_1, v_2\} = \{(2, 1), (1, 1)\}$.

Solution. According to the analysis above we need to determine $s_0 [I]_{S_1}$ and $s_1 [I]_{S_0}$. Of course $s_1 [I]_{S_0} = s_0 [I]_{S_1}^{-1}$. Since the coordinates of the vectors in S_1 are expressed in terms of the basis vectors S_0 we obtain directly

$$s_0 [I]_{S_1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Its inverse is given by

$$s_0 [I]_{S_1}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

Assembling these matrices we have the final matrix converted to the new basis.

$$\begin{aligned} s_1 [A]_{S_1} &= s_1 [I]_{S_0} A s_0 [I]_{S_1} \\ &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 4 \\ -7 & -4 \end{bmatrix} \end{aligned}$$

Example 2.8.2. Consider the same problem as above except that the matrix A is given in the basis S_1 . Find matrix representation of this transformation with respect to the basis S_0 .

Solution. To solve this problem we need to determine $s_0 [A]_{S_0} = s_0 [I]_{S_1} A s_1 [I]_{S_0}$. As we already have these matrices, we determine that

$$\begin{aligned} s_0 [A]_{S_0} &= s_0 [I]_{S_1} A s_1 [I]_{S_0} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} -6 & 13 \\ -4 & 8 \end{bmatrix} \end{aligned}$$

2.9 Appendix A — Solving linear systems

The key to solving linear systems is to reduce the augmented system to RREF and solve the resulting equations. While this may be so, there is an intermediate step that occurs about half way through the computation of the RREF where the reduced matrix achieves an upper triangular form. At this point the solution can be determined directly by back substitution. To clarify the rules on back substitution, suppose that we have the triangular form

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn} & b_n \end{array} \right]$$

Assuming that the diagonal part consists of all nonzero terms, we can solve this system by back substitution. First solve for $x_n = \frac{b_n}{a_{nn}}$. Now inductively solve for the remaining solution coordinates using the formula

$$x_{n-j} = \frac{1}{b_{n-j,n-j}} \left[\sum_{k=0}^{j-1} a_{n-j,n-k} x_{n-k} \right], \quad j = 1, 2, \dots, n-1$$

This inconvenient looking formula can be replaced by

$$x_j = \frac{1}{b_{jj}} \left[\sum_{k=j+1}^n a_{jk} x_k \right], \quad j = n-1, n-2, \dots, 1$$

where the index runs from $j = n-1$ up to $j = 1$. The upshot is that the row reduction process can be halted when a triangular-like form has been attained. This applies as well to nonsingular and non square systems, where the process is stopped when all the leading ones have been identified, entries *below* them have been zeroed out, and all the zero rows are present. The principle reason for using back substitution is to reduce the number of computations required, an important consideration in numerical linear algebra. In the example below we solve a 3×3 nonsingular system.

Example 2.9.1. Solve $Ax = b$ where

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 2 & -1 \\ -1 & 3 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 6 \\ -2 \end{bmatrix}$$

Solution. Find the RREF of $[A \mid b]$. Then solve $Ax = b$.

$$\begin{aligned}
 \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 2 & 2 & -1 & 6 \\ -1 & 3 & 2 & -2 \end{array} \right] & \xrightarrow{\begin{array}{l} -2R_1 + R_2 \\ R_1 + R_3 \end{array}} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & -2 & -1 & 0 \\ 0 & 5 & 2 & 1 \end{array} \right] \\
 & \xrightarrow{-\frac{1}{2}R_2} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 5 & 2 & 1 \end{array} \right] \\
 & \xrightarrow{-5R_2 + R_3} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} & 1 \end{array} \right] \quad (*) \\
 & \xrightarrow{-\frac{1}{2}R_3 + R_2} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array} \right] \\
 & \xrightarrow{-2R_3} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & -2 \end{array} \right] \\
 & \xrightarrow{-2R_2 + R_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array} \right]
 \end{aligned}$$

Hence solving we obtain $x_3 = -2$, $x_2 = 1$, and $x_1 = 1$. This is fine, but there is a faster way to solve this system. Stop the reduction when the system attains a triangular form at $(*)$. From this point solve to obtain $x_3 = -2$. Now back substitute $x_3 = -2$ into the second row (equation) to solve for x_2 . Thus $x_2 = -\frac{1}{2}(-2) = 1$. Finally, back substitute $x_3 = -2$ and $x_2 = 1$ into the first row (equation) to solve for x_1 . Thus $x_1 = 3 - 2(1) = 1$. Sometimes the form $(*)$ is called the row reduced form.

Example 2.9.2. Given the augmented system for $Ax = b$ is in RREF.

$$\left[\begin{array}{cccccc|c} 1 & 2 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Find the solution.

Solution. The leading ones occur in columns 1, 3, and 5. The values in columns 2, 4, and 6 can be taken as free parameters. So, take $x_2 = r$, $x_4 = s$, and $x_6 = t$. Now solving for the other variables we have

$$\begin{aligned}x_1 &= 4 - 2r \\x_3 &= 1 - 2s + t \\x_5 &= 1 - 3t\end{aligned}$$

The solution set is comprised of the vector

$$\begin{aligned}x &= [4 - 2r, r, 1 - 2s + t, s, 1 - t, t]^T \\&= [4, 0, 1, 0, 1, 0]^T + r[-2, 1, 0, 0, 0, 0]^T + s[0, 0, -2, 1, 0, 0]^T + t[0, 0, 1, 0 - 3, 1]^T\end{aligned}$$

for all r , s , and t . We can rewrite this as the set

$$S = \left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -3 \\ 1 \end{bmatrix} \mid r, s, t \in R \text{ or } \mathbb{C} \right\}$$

This representation shows better the connection between the free constants and the component vectors that make up the solution. Note this expression also reveals the solution of the homogeneous solution $Ax = 0$ as the set

$$\left\{ r[-2, 1, 0, 0, 0, 0]^T + s[0, 0, -2, 1, 0, 0]^T + t[0, 0, 1, 0 - 3, 1]^T \mid r, s, t \in R \text{ or } \mathbb{C} \right\}$$

Indeed, this is a full subspace.

Example 2.9.3. The RREF can be used to determine the inverse, as well. Given the matrix $A \in M_n$, the inverse is given by the matrix X for which $AX = I$. In turn with $x_1, \dots, x_n$ representing the columns of X and $e_1, \dots, e_n$ representing the standard vectors we see that $Ax_j = e_j$, $j = 1, 2, \dots, n$. To solve for these vectors, form the augmented matrix $[A \mid e_j]$, $j = 1, 2, \dots, n$ and row reduce as above. A massive short cut to this process is to augment all the standard vectors at one and row reduce the

resulting $n \times 2n$ matrix $[A \mid I]$. If A is invertible, its RREF is the identity. Therefore,

$$[A \mid I] \xrightarrow[\text{operations}]{\text{row}} [I \mid X]$$

and, of course, $A^{-1} = X$. Thus, for

$$A = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 1 & 2 \\ 3 & -2 & -1 \end{bmatrix}$$

we row reduce $[A \mid I]$ as follows

$$[A \mid I] = \left[\begin{array}{ccc|ccc} -2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 3 & -2 & -1 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\text{operations}]{\text{row}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 1 & 2 \\ 0 & 1 & 0 & 7 & 2 & 4 \\ 0 & 0 & 1 & -5 & -1 & -3 \end{array} \right]$$

2.10 Exercises

1. Consider the differential operator $T = 2x \frac{d}{dx}(\cdot) - 4$ acting on the vector space of cubic polynomials, P_3 . Show that T is a linear transformation and find a matrix representation of it. Assume the basis is given by $\{1, x, x^2, x^3\}$.
2. (i) Find matrices A and B , each with positive rank, for which $r(A + B) = r(A) + r(B)$. (ii) Find matrices A and B , each with positive rank, for which $r(A + B) = 0$. (iii) Give a method to find two nonzero matrices A and B for which the sum has any preassigned rank. Of course, the matrix sizes may depend on this value.
3. Find square matrices A and B for which $r(A) = r(B) = 2$ and for which $r(AB) = 0$.
4. Suppose that A is an $m \times n$ matrix and that x is a solution of $Ax = b$ over the prescribed field. Show that every solution of $Ax = b$ have the form $x + x_0$, where x_0 is a solution of $Ax_0 = 0$.

5. Find a matrix $A \in M_n$ of rank $n-1$ for which $r(A^k) = n-k$ for $k \leq n$. Is it possible to begin this process with a matrix $A \in M_n$ of rank n and for which $r(A^k) = n-k+1$ for $k \leq n$?
6. Show that $Ax = b$ has a solution if and only if $y^T b = 0$ if and only if $y^T A = 0$ for some column vector.
7. In R_2 the linear transformation that rotates any vector by θ radians counter clockwise T has matrix representation with respect to the standard basis given by

$$A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

What is the matrix representation with respect to the standard basis of the transformation that rotates any vector by θ radians clockwise? What is the relation between the matrices?

8. Show that if $B, C \in M_n(F)$, where B is symmetric and C is skew-symmetric, then $B = C$ implies that $B = C = 0$.
9. Prove the general formula for the inverse of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.
10. Prove Theorem 2.5.1(a).
11. Prove Theorem 2.5.1(b).
12. Prove that every permutation σ must have an inverse σ^{-1} (i.e. $\sigma^{-1}(\sigma(j)) = j$), and the signs of σ^{-1} and σ are the same.
13. Show that the sign of every transposition is -1 .
14. Prove that $\det A = \sum_{\sigma} (-1)^{\text{sgn}(\sigma)} \left(\prod_i a_{\sigma(i) i} \right)$
15. Prove Proposition 2.5.2 using minors.
16. Suppose that the $n \times n$ matrix A is singular. Show that each column of the adjugate matrix $\hat{A}$ is a solution of $Ax = 0$. (McDuffee, Chapter 3, Theorem 29.)

17. Suppose that A is an $(n-1) \times n$ matrix, and consider the homogeneous system $Ax = 0$ for $x \in R_n$. Define h_i to be the determinant of the $(n-1) \times (n-1)$ matrix formed by removing the i^{th} column of A . Show that the vector $h = (h_1, \dots, h_n)^T$ is a solution to $Ax = 0$. (McDuffee, Chapter 3, Corollary 29.)
18. Show that $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ has no inverse by trying to solve $AB = I$. That is, assume the form

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

multiply the matrices (A and B) together, and then solve for the unknowns a , b , c , and d . (This is not a very efficient way to determine inverses of matrices. Try the same thing for any 3×3 matrix.)

19. Prove that the elementary equation operations do not change the solution set of a linear system.
20. Find the inverses of E_1 , E_2 , and E_3 .
21. Find the matrix representation of linear transformation T that rotates any vector by θ radians counter clockwise (ccw) with respect to the basis $S = \{(2, 1), (1, 1)\}$.
22. Consider R_3 . Suppose that we have angles $\{\theta_i\}_{i=1}^3$ and pairs of coordinate vectors $\{(e_1, e_2), (e_1, e_3), (e_2, e_3)\}$. Let T be the linear transformation that successively rotates a vector in the respective planes $\{(e_{i_1}, e_{i_2})\}_{i=1}^k$ through the respective angles $\{\theta_i\}_{i=1}^k$. Find the matrix representation of T with respect to the standard basis. Prove that it is invertible.
23. Prove Theorem 2.2.4.
24. Prove or disprove the equivalence of the linear systems.

$$\begin{array}{ll} 2x - 3y = -1 & -x + 4y = 3 \\ x + 4y = 5 & x + 2y = 3 \end{array}$$

25. Find basis for the orthocomplement of the subspace of R_3 spanned by the vectors $\{[2, 1, 1]^T, [1, 1, 2]^T\}$.
26. Find basis for the orthocomplement of the subspace of R_3 spanned by the vector $[1, 1, 1]^T$.

27. Consider planar rotations in R_n with respect to the standard bases elements. Prove that there must be $\frac{n(n-1)}{2}$ of them — discounting the particular angle. Display the general representation of any of them. Prove or disprove that any two of them are commutative. That is for two angles $\{\theta_i\}_{i=1}^2$ and pairs of coordinate vectors $\{(e_{i_1}, e_{i_2})\}_{i=1}^k$ the respective counter clockwise rotations are commutative.
28. Suppose that $A \in M_{mk}$, $B \in M_{kn}$ and both have rank k . Show that the rank of AB is k .
29. Suppose that $A \in M_{mk}$ has rank k . Prove that $A_{\text{RREF}} = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$ where I_k is the identity matrix of size k and 0 is the $m-k \times k$ zero matrix.
30. Suppose that $B \in M_{kn}$ has rank k . Prove that $B_{\text{RREF}} = \begin{bmatrix} I_k & 0 \end{bmatrix}$ where I_k is the identity matrix of size k and 0 is the $k \times n-k$ zero matrix.
31. Determine and prove a version of Corollary 2.6.1 for 3×3 blocked matrices, where we assume the diagonal blocks A_{11} is invertible and wish to conclude the result that the rank of A is the equal to the rank of A_{11} .
32. Suppose that we have angles $\{\theta_i\}_{i=1}^k$ and pairs of coordinate vectors $\{(e_{i_1}, e_{i_2})\}_{i=1}^k$. Let T be the linear transformation that successively rotates a vector in the respective planes $\{(e_{i_1}, e_{i_2})\}_{i=1}^k$ through the respective angles $\{\theta_i\}_{i=1}^k$. Prove that the matrix representation of the linear transformation with respect to any basis must be invertible.
33. The super-diagonal of a matrix is the set of elements $a_{i,i+1}$. The sub-diagonal of a matrix is the set of elements $a_{i-1,i}$. A tri-banded matrix is one for which the entries are zero above the super-diagonal and below the sub-diagonal. Suppose that for an $n \times n$ tri-banded matrix T , we have $a_{i-1,i} = a$, $a_{ii} = 0$, and $a_{i,i+1} = c$. Prove the following facts:
- If n is odd $\det A = 0$.
 - If $n = 2m$ is even $\det T = (-1)^m a^m c^m$.
34. For the banded matrix of the previous example, prove the following for the powers T^p of T .

- (a) If p is odd, prove that $(T^p)_{ij} = 0$ if $i + j$ is even.
 (b) If p is even, prove that $(T^p)_{ij} = 0$ if $i + j$ is odd.
35. Consider the vector space $P_2(1, 2)$ with inner product defined by $\langle p, q \rangle = \int_1^2 p(x) q(x) dx$. Find an orthogonal basis of $P_2(1, 2)$. (Hint. Begin with the standard basis $\{1, x, x^2\}$. Apply the Gram-Schmidt procedure.)
36. For what values of a and b is the matrix below singular

$$A = \begin{bmatrix} a & 2 & 1 \\ 2 & 1 & b \\ 1 & a & -2 \end{bmatrix}$$

37. The Vandermonde matrix, defined for a sequence of numbers $\{x_1, \dots, x_n\}$, is given by the $n \times n$ matrix

$$V_n = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

Prove that the determinant is given by

$$\det V_n = \prod_{i>j=1}^n (x_i - x_j)$$

38. In the case the x -values are the integers $\{1, \dots, n\}$, prove that $\det V_n$ is divisible by $\prod_{i=1}^n (i-1)!$. (These numbers are called superfactorials.)
39. Prove that for the weighted functional defined in Remark 2.4.1, it is necessary and sufficient that the weights be strictly positive for it to be an inner product.
40. For what values of a and b is the matrix below singular

$$A = \begin{bmatrix} b & 0 & a & 0 \\ 0 & 0 & b & a \\ 0 & a & 0 & b \\ b & a & 0 & 0 \end{bmatrix}$$

41. Find an orthogonal basis for R_2 from the vectors $\{(1, 2), (2, 1)\}$.
42. Find an orthogonal basis of the subspace of R_3 spanned by $\{(1, 0, 1), (0, 1, -1)\}$
43. Suppose that V is a vector space with an inner product, and $S \subset V$. Show that if S is a basis of V , $S^\perp = \{0\}$.
44. Suppose that V is a vector space with an inner product, and $S \subset V$. Show that if $U = \mathfrak{S}(S)$, then $U^\perp = S^\perp$.
45. Let $A \in M_{mn}(F)$. Show it may not be true that $r(A) = r(A^T A) = r(AA^T)$ unless $F = R$, in which case it is true.
46. If $A \in M_n(C)$ is orthogonal, show that the rows and columns of A are orthogonal.
47. If A is orthogonal then A^m is orthogonal for every positive integer m . (This is a part of Theorem 2.4.3(b).)
48. Consider the polynomial space $P_n[-1, 1]$ with the inner product $\langle p, q \rangle = \int_{-1}^1 p(t)q(t) dt$. Show that every polynomial $p \in P_n$ for which $p(1) = p(-1) = 0$ is orthogonal to its derivative.
49. Consider the polynomial space $P_n[-1, 1]$ with the inner product $\langle p, q \rangle = \int_{-1}^1 p(t)q(t) dt$. Show that the subspace of polynomials in even powers (e.g. $p(t) = t^2 - 5t^6$) is orthogonal to the subspace of polynomials in odd powers.
50. Let $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$ be the matrix representation of a linear transformation given with respect to the standard basis $S_0 = \{e_1, e_2\} = \{(1, 0), (0, 1)\}$ Find the matrix representation of this transformation with respect to the basis $S_1 = \{v_1, v_2\} = \{(2, -3), (1, -2)\}$.
51. Show that the sign of every transposition from the set $\{1, 2, \dots, n\}$ is -1 .
52. What are the signs of the permutations $\{7, 6, 5, 4, 3, 2, 1\}$ and $\{7, 1, 6, 4, 3, 5, 2\}$ of the integers $\{1, 2, 3, 4, 5, 6, 7\}$?
53. Prove that the sign of the permutation $\{m, m-1, \dots, 2, 1\}$ is $(-1)^m$.
54. Suppose $A, B \in M_n(F)$. If A is singular, use a row space argument to show that $\det AB = 0$.

55. Show that if $A \in M_{m,n}$ and $B \in M_{m,n}$ and both $r(A) = r(B) = m$. If $r(AB) = m - k$, what can be said about n ?
56. Prove that

$$\det \begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & -x_2 & x_1 \end{vmatrix} = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^2$$

57. Prove that there is no invertible 3×3 matrix that has all the same cofactors. What similar statement can be made for $n \times n$ matrices?
58. Show by example that there are matrices A and B for which $\lim_{n \rightarrow \infty} A^n$ and $\lim_{n \rightarrow \infty} B^n$ both exist, but for which $\lim_{n \rightarrow \infty} (AB)^n$ does not exist.
59. Let $A \in M_2(\mathbb{C})$. Show that there is no matrix solution $B \in M_2(\mathbb{C})$ to $AB - BA = I$. What can you say about the same problem with $A, B \in M_n(\mathbb{C})$?
60. Show by example that if $AC = BC$ then it does not follow that $A = B$. However, show that if C is invertible the conclusion $A = B$ is valid.

Chapter 3

Eigenvalues and Eigenvectors

In this chapter we begin our study of the most important, and certainly the most dominant aspect, of matrix theory. Called spectral theory, it allows us to give fundamental structure theorems for matrices and to develop power tools for comparing and computing with matrices. We begin with a study of norms on matrices.

3.1 Matrix Norms

We know M_n is a vector space. It is most useful to apply a metric on this vector space. The reasons are manifold, ranging from general information of a metrized system to perturbation theory where the “smallness” of a matrix must be measured. For that reason we define metrics called matrix norms that are regular norms with one additional property pertaining to the matrix product.

Definition 3.1.1. Let $A \in M_n$. Recall that a *norm*, $\|\cdot\|$, on any vector space satisfies the properties:

- (i) $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = 0$
- (ii) $\|cA\| = |c|\|A\|$ for $c \in R$
- (iii) $\|A + B\| \leq \|A\| + \|B\|$.

There is a true vector product on M_n defined by matrix multiplication. In this connection we say that the norm is **submultiplicative** if

- (iv) $\|AB\| \leq \|A\|\|B\|$

In the case that the norm $\|\cdot\|$ satisfies all four properties (i) - (iv) we call it a **matrix norm**.

Here are a few simple consequences for matrix norms. The proofs are straightforward.

Proposition 3.1.1. *Let $\|\cdot\|$ be a matrix norm on M_n , and suppose that $A \in M_n$. Then*

$$(a) \|A\|^2 \leq \|A\|^2, \|A^p\| \leq \|A\|^p, \quad p = 2, 3, \dots$$

$$(b) \text{ If } A^2 = A \text{ then } \|A\| \geq 1$$

$$(c) \text{ If } A \text{ is invertible, then } \|A^{-1}\| \geq \frac{\|I\|}{\|A\|}$$

$$(d) \|I\| \geq 1.$$

Proof. The proof of (a) is a consequence of induction. Supposing that $A^2 = A$, we have by the submultiplicativity property that $\|A\| = \|A^2\| \leq \|A\|^2$. Hence $\|A\| \geq 1$, and therefore (b) follows. If A is invertible, we apply the submultiplicativity again to obtain $\|I\| = \|AA^{-1}\| \leq \|A\| \|A^{-1}\|$, whence (c) follows. Finally, (d) follows because $I^2 = I$ and (b) applies. $\square$

Matrices for which $A^2 = A$ are called **idempotent**. Idempotent matrices turn up in most unlikely places and are useful for applications.

Examples. We can easily apply standard vector space type norms, i.e. ℓ_1 , ℓ_2 , and ℓ_∞ to matrices. Indeed, an $n \times n$ matrix can clearly be viewed as an element of $\mathbb{C}_{n^2}$ with the coordinate stacked in rows of n numbers each. The trick is usually to verify the submultiplicativity condition (iv).

1. ℓ_1 . Define

$$\|A\|_1 = \sum_{i,j} |a_{ij}|$$

The usual norm conditions (i)–(iii) hold. To show submultiplicativity we write

$$\begin{aligned} \|AB\|_1 &= \sum_{ij} \left| \sum_k a_{ik} b_{kj} \right| \\ &\leq \sum_{ij} \sum_k |a_{ik}| |b_{kj}| \\ &\leq \sum_{ijk} |a_{ik}| |b_{mj}| = \sum_{i,k} |a_{ik}| \sum_{mj} |b_{mj}| \\ &= \|A\|_1 \|B\|_1. \end{aligned}$$

Thus $\|A\|_1$ is a matrix norm.

2. ℓ_2 . Define

$$\|A\|_2 = \left(\sum_{i,j} a_{ij}^2 \right)^{1/2}$$

conditions (i)–(iii) clearly hold. $\|A\|_2$ is also a matrix norm as we see by application of the Cauchy–Schwartz inequality. We have

$$\begin{aligned} \|AB\|_2^2 &= \sum_{i,j} \left(\sum_k a_{ik} b_{kj} \right)^2 \leq \sum_{i,j} \left(\sum_k a_{ik}^2 \right) \left(\sum_m b_{jm}^2 \right) \\ &= \left(\sum_{i,k} |a_{ik}|^2 \right) \left(\sum_{j,m} |b_{jm}|^2 \right) \\ &= \|A\|_2^2 \|B\|_2^2. \end{aligned}$$

This norm has three common names: The (a) Frobenius norm, (b) Schur norm, and (c) Hilbert–Schmidt norm. It has considerable importance in matrix theory.

3. ℓ_∞ . Define for $A \in M_n(R)$

$$\|A\|_\infty = \sup_{i,j} |a_{ij}| = \max_{i,j} |a_{ij}|.$$

Note that if $J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\|J\|_\infty = 1$. Also $J^2 = 2J$. Thus $\|J^2\| = 2\|J\| = 1 \not\leq \|J\|^2$. So $\|A\|_\infty$ is not a matrix norm, though it is a vector space norm. We can make it into a matrix norm by

$$\|A\| = n\|A\|_\infty.$$

Note

$$\begin{aligned} |||AB||| &= n \max_{i,j} \left| \sum_k a_{ik} b_{kj} \right| \\ &\leq n \max_{i,j} n \max_k |a_{ik}| |b_{kj}| \\ &\leq n^2 \max_{i,k} |a_{ik}| \max_{k,j} |b_{kj}| \\ &= |||A||| \quad |||B|||. \end{aligned}$$

In the inequalities above we use the fundamental inequality

$$\sum_k |c_k d_k| \leq \max_k |d_k| \sum_k |c_k|$$

(See Exercise 4.) While these norms have some use in general matrix theory, most of the widely applicable norms are those that are subordinate to vector norms in the manner defined below.

Definition 3.1.2. Let $\|\cdot\|$ be a vector norm on R_n (or $\mathbb{C}_n$). For $A \in M_n(R)$ (or $M_n(\mathbb{C})$) we define the norm $\|A\|$ on M_n by

$$\|A\| = \max_{\|x\|=1} \|Ax\|. \quad (\star)$$

and call $\|A\|$ the norm **subordinate** to the vector norm. Note the use of the same notation for both the vector and subordinate norms.

Theorem 3.1.1. *The subordinate norm is a matrix norm and $\|Ax\| \leq \|A\|\|x\|$.*

Proof. We need to verify conditions (i)–(iv). Conditions (i) and (ii) are obvious and are left to the reader. To show (iii), we have

$$\begin{aligned} \|A + B\| &= \max_{\|x\|=1} \|(A + B)x\| \leq \max_{\|x\|=1} (\|Ax\| + \|Bx\|) \\ &\leq \max_{\|x\|=1} \|Ax\| + \max_{\|x\|=1} \|Bx\| \\ &= \|A\| + \|B\|. \end{aligned}$$

Note that

$$\begin{aligned} \|Ax\| &= \|x\| A \left(\frac{x}{\|x\|} \right) \\ &\leq \|A\| \|x\| \end{aligned}$$

since $\left\| \frac{x}{\|x\|} \right\| = 1$. Finally, it follows that for any $x \in R_n$

$$\|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\|$$

and therefore $\|AB\| \leq \|A\| \|B\|$. □

Corollary 3.1.1. *(i) $\|I\| = 1$.*

(ii) If A is invertible, then

$$\|A^{-1}\| \geq (\|A\|)^{-1}.$$

Proof. For (i) we have

$$\|I\| = \max_{\|x\|=1} \|Ix\| = \max_{\|x\|=1} \|x\| = 1.$$

To prove (ii) begin with $A^{-1}A = I$. Then by the submultiplicativity and (i)

$$1 = \|I\| \leq \|A^{-1}\| \|A\|$$

and so $\|A^{-1}\| \geq 1/\|A\|$. □

There are many results connected with matrix norms and eigenvectors that we shall explore before long. The relation between the norm of the matrix and its inverse is important in computational linear algebra. The quantity $\|A^{-1}\| \|A\|$ the **condition number** of the matrix A . When it is very large, the solution of the linear system $Ax = b$ by general methods such as Gaussian elimination may produce results with considerable error. The condition number, therefore, tips off investigators to this possibility. Naturally enough the condition number may be difficult to compute accurately in exactly these circumstances. Alternative and very approximate methods are often used as reliable substitutes for the condition number.

A special type of matrix, one for which $\|Ax\| = \|x\|$ for every $x \in \mathbb{C}$, is called an **isometry**. Such matrices which do not “stretch” any vectors have remarkable spectral properties and play an important roll in spectral theory.

3.2 Convergence and perturbation theory

It will often be necessary to compare one matrix with another matrix that is *nearby* in some sense. When a matrix norm at hand it is possible to measure the proximity of two matrices by computing the norm of their difference. This is just as we do for numbers. We begin with this study by showing that if the norm of a matrix is less than one, then its difference with the identity is invertible. Again, this is just as with numbers; that is, if $|r| < 1$, $\frac{1}{1-r}$ is defined. Let us assume $R \in M_n$ and $\|\cdot\|$ is some norm on M_n . We want to show that if $\|R\| < 1$ then $(I - R)^{-1}$ exists. Toward this end we prove the following lemma.

Lemma 3.2.1. *For every $R \in M_n$*

$$(I - R)(I + R + R^2 + \cdots + R_n) = I - R^{n+1}.$$

Proof. This result for matrices is the direct analog of the result for numbers $(1-r)(1+r+r^2+\cdots+r^n) = 1-r^{n+1}$, also often written as $1+r+r^2+\cdots+r^n = \frac{1-r^{n+1}}{1-r}$. We prove the result inductively. If $n = 1$ the result follows from direct computation, $(I - R)(I + R) = I - R^2$. Assume the result holds up to $n - 1$. Then

$$\begin{aligned} (I - R)(I + R + R^2 + \cdots + R^{n-1} + R^n) &= (I - R)(I + R + \cdots + R^{n-1}) \\ &\quad + (I - R)R^n \\ &= (I - R^n) + R^n - R^{n+1} = I - R^{n+1} \end{aligned}$$

by our inductive hypothesis. This calculation completes the induction, and hence the proof. $\square$

Remark 3.2.1. Sometimes the proof is presented in a “quasi-inductive” manner. That is, you will see

$$\begin{aligned} (I - R)(I + R + R^2 + \cdots + R^n) &= (I + R + R^2 + \cdots + R^n) \\ &\quad - (R + R^2 + \cdots + R^{n+1}) \quad (*) \\ &= I - R^{n+1} \end{aligned}$$

This is usually considered acceptable because the correct induction is transparent in the calculation.

Below we will show that if $\|R\| = \lambda < 1$, then $(I + R + R^2 + \cdots) = (I - R)^{-1}$. It would be *incorrect* to apply the obvious fact that $\|R^{n+1}\| < \lambda^{n+1} \rightarrow 0$ to draw the conclusion from the equality (*) above without first establishing convergence of the series $\sum_0^\infty R^k$. A crucial step in showing that an infinite series is convergent is showing that its partial sums satisfy the Cauchy criterion: $\sum_{k=1}^\infty a_k$ converges if and only if for each $\varepsilon > 0$, there exists an integer N such that if $m, n > N$, then $|\sum_{k=m+1}^n a_k| < \varepsilon$. (See Appendix A.) There is just one more aspect of this problem. While it is easy to establish the Cauchy criterion for our present situation, we still need to resolve the situation between *norm convergence* and *pointwise convergence*. We need to conclude that if $\|R\| < 1$ then $\lim_{n \rightarrow \infty} R^n = 0$, and by this expression we mean that $(R^n)_{ij} \rightarrow 0$ for all $1 \leq i, j \leq n$.

Lemma 3.2.2. *Suppose that the norm $\|\cdot\|$ is a subordinate norm on M_n and $R \in M_n$.*

(i) *If $\|R\| < \varepsilon$, then there is a constant M such that $|p_{ij}| < M\varepsilon$.*

(ii) *If $\lim_{n \rightarrow \infty} \|R^n\| = 0$, then $\lim_{n \rightarrow \infty} R^n = 0$.*

Proof. (i) If $\|R\| < \varepsilon$, it follows that $\|Rx\| < \varepsilon$ for each vector x , and by selecting the standard vectors e_j in turn, it follows that from which it follows that $\|r_{*j}\| < \varepsilon$, where r_{*j} denotes the j^{th} column of R . By Theorem 1.6.2 all norms are equivalent. It follows that there is a fixed constant M independent of ε and R such that $|r_{ij}| < M\varepsilon$.

(ii) Suppose for some increasing subsequence of powers $n_k \rightarrow \infty$ it happens that $\|(R^{n_k})_{ij}\| \geq r$. Select the standard unit vector e_j . A little computation shows that $\|R^{n_k}e_j\| \geq r$, whence $\|R^{n_k}\| \geq r$, contradicting the known limit $\lim_{n \rightarrow \infty} \|R^n\| = 0$. The conclusion $\lim_{n \rightarrow \infty} R^n = 0$ follows. $\square$

Lemma 3.2.3. *Suppose that the norm $\|\cdot\|$ is a subordinate norm on M_n . If $\|R\| = \lambda < 1$, then $I + R + R^2 + \cdots + R^k + \cdots$ converges.*

Proof. Let $P_n = I + R + \cdots + R^n$. To show convergence we establish that $\{P_n\}$ is a Cauchy sequence. For $n > m$ we have

$$P_n - P_m = \sum_{k=m+1}^n R^k$$

Hence

$$\begin{aligned} \|P_n - P_m\| &= \left\| \sum_{k=m+1}^n R^k \right\| \\ &\leq \sum_{k=m+1}^n \|R^k\| \\ &\leq \sum_{k=m+1}^n \|R\|^k \\ &= \sum_{k=m+1}^n \lambda^k = \lambda^{m+1} \sum_{j=0}^{n-m-1} \lambda^j \\ &\leq \lambda^{m+1} (1 - \lambda)^{-1} \rightarrow 0 \end{aligned}$$

where in the second last step we used the inequality, which is valid for $0 \leq \lambda \leq 1$. $\sum_0^{n-m-1} \lambda^j \leq \sum_0^{\infty} \lambda^j < (1 - \lambda)^{-1}$. We conclude by Lemma 3.2.2 that the individual matrix entries of the partial sums converge and thus the series itself converges. $\square$

Note that this result is independent of the particular norm. In practice it is often necessary to select a convenient norm to actually carry out or verify particular computations are valid. In the theorem below we complete the analysis of the matrix version of the geometric series, stating that when the norm of a matrix is less than one, the geometric series based on that matrix converges and the inverse of the difference with the identity exists.

Theorem 3.2.1. *If $R \in M_n(F)$ and $\|R\| < 1$ for some norm, then $(I - R)^{-1}$ exists and*

$$(I - R)^{-1} = I + R + R^2 + \cdots = \sum_{k=0}^{\infty} R^k.$$

Proof. Apply the two previous lemmas. $\square$

The perturbation result alluded to above can now be stated and easily proved. In words this result states that if we begin with an invertible matrix and additively perturb it by a sufficiently small amount the result remains invertible. Overall, this is the first of a series of results where what is proved is that some property of a matrix is preserved under additive perturbations.

Corollary 3.2.1. *If $A, B \in M_n$ and A is invertible, then $A + \lambda B$ is invertible for sufficiently small $|\lambda|$ (in R or $\mathbb{C}$).*

Proof. As above we assume that $\|\cdot\|$ is a norm on $M_n(F)$. It is an easy computation to see that

$$A + \lambda B = A(I + \lambda A^{-1}B).$$

Select λ sufficiently small so that $\|\lambda A^{-1}B\| = |\lambda| \|A^{-1}B\| < 1$. Then by the theorem above, $I + \lambda A^{-1}B$ is invertible. Therefore

$$(A + \lambda B)^{-1} = (I + \lambda A^{-1}B)^{-1} A^{-1}$$

and the result follows. $\square$

Another way of stating this is to say that if $A, B \in M_n$ and A has a nonzero determinant, then for sufficiently small λ the matrix $A + \lambda B$ also has a nonzero determinant. This corollary can be applied directly to the identity matrix itself being perturbed by a rank one matrix. In this case the λ can be specified in terms of the two vectors comprising the matrix. (Recall Theorem 2.3.1(8).)

Corollary 3.2.2. *Let $x, y \in R_n$ satisfy $|\langle x, y \rangle| = |\lambda| < 1$. Then $I + xy^T$ is invertible and*

$$(I + xy^T)^{-1} = I - xy^T(1 + \lambda)^{-1}.$$

Proof. We have that $(I + xy^T)^{-1}$ exists by selecting a norm $\|\cdot\|$ consistent with the inner product $\langle \cdot, \cdot \rangle$. (For example, take $\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2$, where $\|\cdot\|_2$ is the Euclidean norm.) It is easy to see that $(xy^T)^k = \lambda^{k-1}xy^T$. Therefore

$$\begin{aligned} (I + xy^T)^{-1} &= I - xy^T + (xy^T)^2 - (xy^T)^3 + \cdots \\ &= I - xy^T + \lambda xy^T - \lambda^2 xy^T + \cdots \\ &= I - xy^T \left(\sum_{k=0}^{\infty} (-\lambda)^k \right). \end{aligned}$$

Thus

$$(I + xy^T)^{-1} = I - xy^T(1 + \lambda)^{-1}$$

and the result is proved. $\square$

In words we conclude that the perturbation of the identity by a small rank 1 matrix has a *computable* inverse.

3.3 Eigenvectors and Eigenvalues

Throughout this section we will consider only matrices $A \in M_n(\mathbb{C})$ or $M_n(R)$. Furthermore, we suppress the field designation unless it is relevant.

Definition 3.3.1. If $A \in M_n$ and $x \in \mathbb{C}_n$ or R_n . If there is a constant $\lambda \in \mathbb{C}$ and a vector $x \neq 0$ for which

$$Ax = \lambda x$$

we call λ an **eigenvalue** of A and x its corresponding **eigenvector**. Alternatively, we call x the eigenvector *pertaining* to the eigenvalue λ , and vice-versa.

Definition 3.3.2. For $A \in M_n$, define

- (1) $\sigma(A) = \{\lambda \mid Ax = \lambda x \text{ has a solution for a nonzero vector } x\}$. $\sigma(A)$ is called the **spectrum** of A .
- (2) $\rho(A) = \sup_{\lambda \in \sigma(A)} |\lambda|$ (or equivalently $\max_{\lambda \in \sigma(A)} |\lambda|$). $\rho(A)$ is called the **spectral radius**.

Example 3.3.1. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Then $\lambda = 1$ is an eigenvalue of A with eigenvector $x = [-1, 1]^T$. Also $\lambda = 3$ is an eigenvalue of A with eigenvector $x = (1, 1)^T$. The spectrum of A is $\sigma(A) = \{1, 3\}$ and the spectral radius of A is $\rho(A) = 3$.

Example 3.3.2. The 3×3 matrix $B = \begin{bmatrix} -3 & 0 & 6 \\ -12 & 9 & 26 \\ 4 & -4 & -9 \end{bmatrix}$ has eigenvalues:

$-1, -3, 1$. Pertaining to the eigenvalues are the eigenvectors

$$\left\{ \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \right\} \leftrightarrow 1, \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\} \leftrightarrow -3, \left\{ \begin{bmatrix} 3 \\ -2 \\ 2 \end{bmatrix} \right\} \leftrightarrow -1$$

The characteristic polynomial

To say that $Ax = \lambda x$ has a nontrivial solution ($x \neq 0$) for some $\lambda \in \mathbb{C}$ is the same as the assertion that $(A - \lambda I)x = 0$ has a nontrivial solution. This means that

$$\det(A - \lambda I) = 0$$

or what is more commonly written

$$\det(\lambda I - A) = 0.$$

From the original definition (Definition 2.5.1) the determinant is sum of products of individual matrix entries. Therefore, $\det(\lambda I - A)$ must be a polynomial in λ . This makes the definition:

Definition 3.3.3. Let $A \in M_n$. The determinant

$$p_A(\lambda) = \det(\lambda I - A)$$

is called the **characteristic polynomial** of A . Its zeros¹ are called the **eigenvalues** of A . The set $\sigma(A)$ of all eigenvalues of A is called the **spectrum** of A .

A simple consequence of the nature of the determinant of $\det(\lambda I - A)$ is the following.

Proposition 3.3.1. *If $A \in M_n$, then $p_A(\lambda)$ has degree exactly n .*

See Appendix A for basic information on solving polynomials equations $p(\lambda) = 0$. We may note that even though $A \in M_n(\mathbb{C})$ has n^2 entries in its definition, its spectrum is completely determined by the n coefficients of $p_A(\lambda)$.

Procedure. The basic procedure of determining eigenvalues and eigenvectors is this: (1) Solve $\det(\lambda I - A) = 0$ for the eigenvalues λ and (2) for any given eigenvalue λ solve the system $(A - \lambda I)x = 0$ for the pertaining eigenvector(s). Though this procedure is not practical in general it can be effective for small sized matrices, and for matrices with special structures.

Theorem 3.3.1. *Let $A \in M_n$. The set of eigenvectors pertaining to any particular eigenvalue is a subspace of the given vector space $\mathbb{C}_n$.*

Proof. Let $\lambda \in \sigma(A)$. The set of eigenvectors pertaining to λ is the null space of $(A - \lambda I)$. The proof is complete by application of Theorem 2.3.1 that states the null space of any matrix is a subspace of the underlying vector space. $\square$

In light of this theorem the following definition makes sense and is a most important concept in the study of eigenvalues and eigenvectors.

Definition 3.3.4. Let $A \in M_n$ and let $\lambda \in \sigma(A)$. The null space of $(A - \lambda I)$ is called the **eigenspace** of A pertaining to λ .

Theorem 3.3.2. *Let $A \in M_n$. Then*

- (i) *Eigenvectors pertaining to different eigenvalues are linearly independent.*

¹The zeros of a polynomial (or more generally a function) $p(\lambda)$ are the solutions to the equation $p(\lambda) = 0$. A solution to $p(\lambda) = 0$ is also called a *root* of the equation.

(ii) Suppose $\lambda \in \sigma(A)$ with eigenvector x is different from the set of eigenvalues $\{\mu_1, \dots, \mu_k\} \subset \sigma(A)$ and V_μ is the span of the pertaining eigenspaces. Then $x \notin V_\mu$.

Proof. (i) Let $\mu, \lambda \in \sigma(A)$ with $\mu \neq \lambda$ pertaining eigenvectors x and y respectively. Suppose these vectors are linearly dependent; that is, $y = cx$. Then

$$\begin{aligned}\mu y &= \mu cx = c\mu x \\ &= cAx = A(cx) \\ &= Ay \\ &= \lambda y\end{aligned}$$

This is a contradiction, and (i) is proved.

(ii) Suppose the contrary holds, namely that $x \in V_\mu$. Then $x = a_1y_1 + \dots + a_ky_m$ where $\{y_1, \dots, y_m\}$ are linearly independent vectors of V_μ . Each of the vectors y_i is an eigenvector, we know. Assume the pertaining eigenvalues denoted by μ_{j_i} . That is to say, $Ay_i = \mu_{j_i}y_i$, for each $i = 1, \dots, m$. Then

$$\begin{aligned}\lambda x &= Ax = A(a_1y_1 + \dots + a_my_m) \\ &= a_1\mu_{j_1}y_1 + \dots + a_k\mu_{j_m}y_m\end{aligned}$$

If $\lambda \neq 0$ we have

$$x = a_1y_1 + \dots + a_my_m = a_1\frac{\mu_{j_1}}{\lambda}y_1 + \dots + a_k\frac{\mu_{j_m}}{\lambda}y_m$$

We know by the previous part of this result that at least two of the coefficients a_i must be nonzero. (Why?) Thus we have two different representations of the same vector by linearly independent vectors, which is impossible. On the other hand, if $\lambda = 0$ then $a_1\mu_{j_1}y_1 + \dots + a_k\mu_{j_m}y_k = 0$, which is also impossible. Thus, (ii) is proved. $\square$

The examples below will illustrate the spectra of various matrices.

Example 3.3.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then $p_A(\lambda)$ is given by

$$\begin{aligned}p_A(\lambda) &= \det(\lambda I - A) = \det \begin{bmatrix} \lambda - a & -b \\ -c & \lambda - d \end{bmatrix} \\ &= (\lambda - a)(\lambda - d) - bc \\ &= \lambda^2 - (a + d)\lambda + ad - bc.\end{aligned}$$

The eigenvalues are the roots of $p_A(\lambda) = 0$

$$\begin{aligned}\lambda &= \frac{a + d \pm \sqrt{(a + d)^2 - 4(ad - bc)}}{2} \\ &= \frac{a + d \pm \sqrt{(a - d)^2 + 4bc}}{2}.\end{aligned}$$

For this quadratic there are three possibilities:

- (a) Two real roots $\begin{cases} \nearrow \text{different values} \\ \searrow \text{equal values} \end{cases}$
- (b) Two complex roots

Here are three 2×2 examples that illustrate each possibility. The reader should compute the characteristic polynomials to verify these computation.

$$B_1 = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}. \quad \text{Then } p_{B_1}(\lambda) = \lambda^2 \quad \lambda = 0, 0$$

$$B_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad \text{Then } p_{B_2}(\lambda) = \lambda^2 + 1 \quad \lambda = \pm i$$

$$B_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad \text{Then } p_{B_3}(\lambda) = \lambda^2 - 1 \quad \lambda = \pm 1.$$

Example 3.3.2. Consider the rotation in the x, z -plane through an angle θ

$$B = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}$$

The characteristic polynomial is given by

$$\begin{aligned}p_B(\lambda) &= \det \left(\lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \right) \\ &= -1 + (1 + 2 \cos \theta) \lambda - (1 + 2 \cos \theta) \lambda^2 + \lambda^3\end{aligned}$$

The eigenvalues are $1, \cos \theta + \sqrt{\cos^2 \theta - 1}, \cos \theta - \sqrt{\cos^2 \theta - 1}$. When θ is not equal to an even multiple of π , exactly two of the roots are complex numbers. In fact, they are complex conjugate pairs, which can also be written as $\cos \theta + i \sin \theta, \cos \theta - i \sin \theta$. The magnitude of each eigenvalue

is 1, which means all three eigenvalues lie on the unit circle in the complex plane. An interesting observation is that the characteristic polynomial and hence the eigenvalues are the same regardless of which pair of axes (x - z , x - y , or y - z) is selected for the rotation. Matrices of the form B are actually called rotations. In two dimensions the counter-clockwise rotations through the angle θ are given by

$$B_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

The eigenvalues for all θ is not equal to an even multiple of π are $\pm i$. (See Exercise 2.)

Example 3.3.3. If T is upper triangular with $\text{diag } T = [t_{11}, t_{22}, \dots, t_{nn}]$. Then $\lambda I - T$ is upper triangular with $\text{diag}[\lambda - t_{11}, \lambda - t_{22}, \dots, \lambda - t_{nn}]$. Thus the determinant of $\lambda I - T$ gives the characteristic polynomial of T to be

$$p_T(\lambda) = \prod_{i=1}^n (\lambda - t_{ii})$$

The eigenvalues of T are the diagonal elements of T . By expanding this product we see that

$$p_T(\lambda) = \lambda^n - (\sum t_{ii})\lambda^{n-1} + \text{lower order terms.}$$

The constant term of $p_T(\lambda)$ is $(-1)^n \prod_{i=1}^n t_{ii} = (-1)^n \det T$. We define

$$\text{tr } T = \sum_{i=1}^n t_{ii}$$

and call it the **trace** of T . The same statements apply to lower triangular matrices. Moreover, the trace definition applies to all matrices, not just to triangular ones, and the result will be the same.

Example 3.3.4. Suppose that A is rank 1. Then there are two vectors $w, z \in \mathbb{C}_n$ for which

$$A = wz^T.$$

To find the spectrum of A we consider the equation

$$Ax = \lambda x$$

or

$$\langle z, x \rangle w = \lambda x.$$

From this we see that $x = w$ is an eigenvector with eigenvalue $\langle z, w \rangle$. If $z \perp x$, x is an eigenvector pertaining to the eigenvalue 0. Therefore,

$$\sigma_A = \{\langle z, w \rangle, 0\}.$$

The characteristic polynomial is

$$p_A(\lambda) = (\lambda - \langle z, w \rangle)\lambda^{n-1}.$$

If w and z are orthogonal then

$$p_A(\lambda) = \lambda^n.$$

If w and z are not orthogonal though there are just two eigenvalues, we say that 0 is an eigenvalue of *multiplicity* $n - 1$, the order of the factor $(\lambda - 0)$. Also $\langle z, w \rangle$ has multiplicity 1. This is the subject of the next section.

For instance, suppose $w = (1, -1, 2)$ and $z = (0, 1, -3)$. Then spectrum of the matrix $A = wz^T$ is given by $\sigma(A) = \{-7, 0\}$. The eigenvalue pertaining to $\lambda = -7$ is w and we may take $x = (0, 3, 1)$ and $(c, 3, 1)$, for any $c \in \mathbb{C}$, to be eigenvectors pertaining to $\lambda = 0$. Note that $\{w, (0, 3, 1), (1, 3, 1)\}$ form a basis for R_3 .

To complete our discussion of characteristic polynomials we prove a result that every n^{th} degree polynomial with lead coefficient one is the characteristic polynomial of some matrix. You will note the similarity of this result and the analogous result for differential systems.

Theorem 3.3.3. *Every polynomial of n^{th} degree with lead coefficient 1, that is*

$$q(\lambda) = \lambda^n + b_1\lambda^{n-1} + \cdots + b_{n-1}\lambda + b_n$$

is the characteristic polynomial of some matrix.

Proof. We consider the $n \times n$ matrix

$$B = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & \ddots & \\ -b_n & -b_{n-1} & & & -b_1 \end{bmatrix}$$

Then $\lambda I - B$ has the form

$$\lambda I - B = \begin{bmatrix} \lambda & -1 & 0 & \cdots & 0 \\ 0 & \lambda & -1 & & 0 \\ \vdots & & \lambda & \ddots & \\ b_n & b_{n-1} & & & \lambda + b_1 \end{bmatrix}$$

Now expand in minors across the bottom row to get

$$\begin{aligned} \det(\lambda I - B) &= b_n (-1)^{n+1} \det \begin{bmatrix} -1 & 0 & \cdots & 0 \\ \lambda & -1 & & 0 \\ & \lambda & \ddots & \end{bmatrix} \\ &\quad + b_{n-1} (-1)^{n+2} \det \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & -1 & & 0 \\ \vdots & \lambda & \ddots & \end{bmatrix} + \cdots \\ &\quad + b_1 (-1)^{n+n} \det \begin{bmatrix} \lambda & -1 & 0 & \cdots \\ 0 & \lambda & -1 & \\ \vdots & & \lambda & \ddots \end{bmatrix} \\ &= b_n (-1)^{n+1} (-1)^{n-1} + b_{n-1} (-1)^{n+2} \lambda (-1)^{n-2} + \cdots \\ &\quad + (\lambda + b_1) (-1)^{n+n} \lambda^{n-1} \\ &= b_n + b_{n-1} \lambda + \cdots + b_1 \lambda^{n-1} + \lambda^n \end{aligned}$$

which is what we set out to prove. (The reader should check carefully the term with b_{n-2} to fully understand the nature of this proof.) $\square$

Multiplicity

Let $A \in M_n(\mathbb{C})$. Since $p_A(\lambda)$ is a polynomial of degree exactly n , it must have exactly n eigenvectors (i.e. roots) $\lambda_1, \lambda_2, \dots, \lambda_n$ counted according to multiplicity. Recall that the multiplicity of an eigenvalue is the number of times the monomial $(\lambda - \lambda_i)$ is repeated in the factorization of $p_A(\lambda)$. For example the multiplicity of the root 2 in the polynomial $(\lambda - 2)^3(\lambda - 5)$ is 3. Suppose $\mu_1, \dots, \mu_k$ are the distinct eigenvalues with multiplicities $m_1, m_2, \dots, m_k$ respectively. Then the characteristic polynomial can be

rewritten as

$$\begin{aligned} p_A(\lambda) = \det(\lambda I - A) &= \prod_{i=1}^n (\lambda - \lambda_i) \\ &= \prod_{i=1}^k (\lambda - \mu_i)^{m_i}. \end{aligned}$$

More precisely, the multiplicities $m_1, m_2, \dots, m_k$ are called the *algebraic multiplicities* of the respective eigenvalues. This factorization will be very useful later.

We know that for each eigenvalue λ of any multiplicity m , there must be at least *one* eigenvector pertaining to λ . What is desired, but not always possible, is to find μ linearly independent eigenvectors corresponding to the eigenvalue λ of multiplicity m . This state of affairs makes matrix theory at once much more challenging but also much more interesting.

Example 3.3.3. For the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{7}{4} & \frac{1}{4}\sqrt{3} \\ 0 & \frac{1}{4}\sqrt{3} & \frac{5}{4} \end{bmatrix}$$

the characteristic polynomial is $\det(\lambda I - A) = p_A(\lambda) = \lambda^3 - 5\lambda^2 + 8\lambda - 4$, which can be factored as $p_A(\lambda) = (\lambda - 1)(\lambda - 2)^2$. We see that the eigenvalues are 1, 2, and 2. So, the multiplicity of the eigenvalue $\lambda = 1$ is 1, and the multiplicity of the eigenvalue $\lambda = 2$ is 2. The eigenspace

pertaining to the eigenvalue $\lambda = 1$ is generated by the vector $\begin{bmatrix} 0 \\ -\frac{1}{3}\sqrt{3} \\ 1 \end{bmatrix}$,

and the dimension of this eigenspace is one. The eigenspace pertaining to

$\lambda = 2$ is generated by $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ \frac{1}{3}\sqrt{3} \end{bmatrix}$. (That is, these two vectors form

a basis of the eigenspace.) To summarize, for the given matrix there is one eigenvector for the eigenvalue $\lambda = 1$, and there are two linearly independent eigenvectors for the eigenvalue $\lambda = 2$. The dimension of the eigenspace pertaining to $\lambda = 1$ is one, and the dimension of the eigenspace pertaining to $\lambda = 2$ is two.

Now contrast the above example where the eigenvectors span the space $\mathbb{C}_3$ and the next example where we have an eigenvalue of multiplicity three but the eigenspace is of dimension one.

Example 3.3.4. Consider the matrix

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

The characteristic polynomial is given by $(\lambda - 2)^3$. Hence the eigenvalue $\lambda = 2$ has multiplicity three. The eigenspace pertaining to $\lambda = 2$ is generated by the single vector $[0, 0, 1]^T$. To see this we solve

$$\begin{aligned} (A - 2I)x &= \left(\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) x \\ &= \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} x = 0 \end{aligned}$$

The row reduced echelon form for $\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. From this it is apparent that we may take $x_3 = t$, but that $x_1 = x_2 = 0$. Now assign $t = 1$ to obtain the generating vector $[0, 0, 1]^T$. This type of example and its consequences seriously complexifies the study of matrices.

Symmetric Functions

Definition 3.3.5. Let n be a positive integer and $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ be given numbers. Suppose that k is a positive integer with $1 \leq k \leq n$. The k^{th} **elementary symmetric function** on the Λ is defined by

$$S_k(\lambda_1, \dots, \lambda_n) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \prod_{j=1}^k \lambda_{i_j}.$$

It is easy to see that

$$\begin{aligned} S_1(\lambda_1, \dots, \lambda_n) &= \sum_{i=1}^n \lambda_i \\ S_n(\lambda_1, \dots, \lambda_n) &= \prod_{i=1}^n \lambda_i. \end{aligned}$$

For a given matrix $A \in M_n$ there are n symmetric functions defined with respect to its eigenvalues $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$. The symmetric functions are sometimes called the invariants of matrices as they are invariant under similarity transformations that will be in Section 3.5. They also furnish directly the coefficients of the characteristic polynomial. Thus specifying the n symmetric functions of an $n \times n$ matrix is sufficient to determine its eigenvalues.

Theorem 3.3.4. *Let $A \in M_n$ have symmetric functions S_k , $k = 1, 2, \dots, n$. Then*

$$\det(\lambda I - A) = \prod_{i=1}^n (\lambda - \lambda_i) = \lambda^n + \sum_{k=1}^n (-1)^k S_k \lambda^{n-k}.$$

Proof. The proof is a consequence of actually expanding the product $\prod_{i=1}^n (\lambda - \lambda_i)$. Each term in the expansion has exactly n terms multiplied together that are combinations of the factor λ and the $-\lambda'_i$'s. For example, for the power λ^{n-k} the coefficient is obtained by computing the total number of products of k "different" $-\lambda'_i$'s. (The term different is in quotes because it refers to different indices not actual values.) Collecting all these terms is accomplished by addition. Now the number of ways we can obtain products of these k different $(-1)\lambda'_i$'s is easily seen to be the number of sequences in the set $\{1 \leq i_1 < \dots < i_k \leq n\}$. The sum of these is clearly $(-1)^k S_k$, with the $(-1)^k$ factor being the collected product of k -1 's. $\square$

Two of the symmetric functions are familiar. In the following we restate this using familiar terms.

Theorem 3.3.5. *Let $A \in M_n(\mathbb{C})$. Then*

$$p_A(\lambda) = \det(\lambda I - A) = \sum_{k=0}^n p_k \lambda^{n-k}$$

where $p_0 = 1$ and

$$\begin{aligned} (i) \quad p_1 &= -\text{tr}(A) = -\sum_{i=1}^n a_{ii} \\ (ii) \quad p_n &= -\det A. \\ (iii) \quad p_k &= (-1)^k S_k, \text{ for } 1 < k < n. \end{aligned}$$

Proof. Note that $p_A(0) = -\det(A) = p_n$. This gives (ii). To establish (i), we consider

$$\det \begin{bmatrix} \lambda \cdot a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & & -a_{2n} \\ \vdots & & \ddots & \\ -a_{n1} & \cdots & & \lambda - a_{nn} \end{bmatrix}.$$

Clearly the product $\prod_1^n (\lambda - a_{ii})$ is one of the selections of products in the calculation process. In every other product there must be no more than $n - 2$ diagonal terms. Hence

$$p_A(\lambda) = \det(\lambda I - A) = \prod_1^n (\lambda - a_{ii}) + p_{n-2}(\lambda),$$

where $p_{n-2}(\lambda)$ is a polynomial of degree $n - 2$. The coefficient of λ^{n-1} is $-\sum_1^n a_{ii}$ by the Theorem 3.3.4, and this is (i). $\square$

As a final note, observe that the characteristic polynomial is defined by knowing the n symmetric functions. However, the matrix itself has n^2 entries. Therefore, one may expect that knowing the only characteristic polynomial of a matrix is insufficient to characterize it. This is correct. Many matrices having rather different properties can have the same characteristic polynomial.

3.4 The Hamilton-Cayley Theorem

The Hamilton-Cayley Theorem opens the doors to a finer analysis of a matrix through the use of polynomials, which in turn is an important tool of spectral analysis. The results states that any square matrix satisfies its own characteritic polynomial, that is $p_A(A) = 0$. The proof is not difficult, but we need some preliminary results about factoring matrix-valued polynomials. Preceding that we need to consider matrix polynomials in some detail.

Matrix polynomials

One of the very important results of matrix theory is the Hamilton-Cayley theorem which states that a matrix satisfies its only characteristic equation. This implies we need the notion of a matrix polynomial. It is an easy idea

— just replace the coefficients of any polynomial by matrices — but it bears some important consequences.

Definition 3.4.1. Let $A_0, A_1, \dots, A_m$ be square $n \times n$ matrices. We can define the polynomial with matrix coefficients

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \dots + A_{m-1}\lambda + A_m.$$

The *degree* of $A(\lambda)$ is m , provided $A_0 \neq 0$. $A(\lambda)$ is called *regular* if $\det A_0 \neq 0$. In this case we can construct an equivalent monic² polynomial.

$$\begin{aligned}\tilde{A}(\lambda) &= A_0^{-1}A(\lambda) = A_0^{-1}A_0\lambda^m + A_0^{-1}A_1\lambda^{m-1} + \dots + A_0^{-1}A_m \\ &= I\lambda^m + \tilde{A}_1\lambda^{m-1} + \dots + \tilde{A}_m.\end{aligned}$$

The algebra of matrix polynomials mimics the normal polynomial algebra. Let

$$A(\lambda) = \sum_0^m A_i\lambda^{m-i} \quad B(\lambda) = \sum_0^m B_i\lambda^{m-i}.$$

(1) **Addition:**

$$A(\lambda) \pm B(\lambda) = \sum_0^m (A_i \pm B_i)\lambda^{m-i}.$$

(2) **Multiplication:**

$$A(\lambda)B(\lambda) = \sum_{i=0}^m \lambda^{m-i} \left(\sum_{k=0}^i A_k B_{m-i+k} \right)$$

The term $\sum_{k=0}^i A_k B_{m-i+k}$ is called the *Cauchy product* of the sequences. Note that the matrices A_j always multiply on the left of the B_k .

(3) **Division:** Let $A(\lambda)$ and $B(\lambda)$ be two matrix polynomials of degree m (as above) and suppose $B(\lambda)$ is regular, i.e. $\det B_0 \neq 0$. We say that $Q_r(\lambda)$ and $R_r(\lambda)$ are **right quotient** and **remainder** of $A(\lambda)$ upon division by $B(\lambda)$ if

$$A(\lambda) = Q_r(\lambda)B(\lambda) + R_r(\lambda) \tag{1}$$

²Recall that a monic polynomial is a polynomial where coefficient of the highest power is one. For matrix polynomials the corresponding coefficient is I , the identity matrix.

if the degree of $R_r(\lambda)$ is less than that of $B(\lambda)$. Similarly $Q_\ell(\lambda)$ and $R_\ell(\lambda)$ are respectively the **left quotient** and **remainder** of $A(\lambda)$ upon division by $B(\lambda)$ if

$$A(\lambda) = B(\lambda)Q_\ell(\lambda) + R_\ell(\lambda) \quad (2)$$

if the degree of $R_\ell(\lambda)$ is less than that of $B(\lambda)$.

In the case (1) we see

$$A(\lambda)B^{-1}(\lambda) = Q_r(\lambda) + R_r(\lambda)B^{-1}(\lambda), \quad (3)$$

which looks much like $\frac{a}{b} = q + \frac{r}{b}$, a way to write the quotient and remainder of a divided by b when a and b are numbers. Also, the form (3) may not properly exist for all λ .

Lemma 3.4.1. *Let $B_i \in M_n(\mathbb{C})$, $i = 0, \dots, n$ with B_0 nonsingular.*

Then the polynomial

$$B(\lambda) = B_0\lambda^n + B_1\lambda^{n-1} + \dots + B_n$$

is invertible for sufficiently large $|\lambda|$.

Proof. We factor $B(\lambda)$ as

$$\begin{aligned} B(\lambda) &= B_0\lambda^n (I + B_0^{-1}B_1\lambda^{-1} + \dots + B_0^{-1}B_n\lambda^{-n}) \\ &= B_0\lambda^n [I + \lambda^{-1}(B_0^{-1}B_1 + \dots + B_0^{-1}B_n\lambda^{1-n})] \end{aligned}$$

For $\lambda > 1$, the norm of the term $B_0^{-1}B_1 + \dots + B_0^{-1}B_n\lambda^{1-n}$ is bounded by

$$\begin{aligned} \|B_0^{-1}B_1 + \dots + B_0^{-1}B_n\lambda^{1-n}\| &\leq \|B_0^{-1}B_1\| + \|B_0^{-1}B_2\| |\lambda|^{-1} + \dots + \|B_0^{-1}B_n\| |\lambda|^{1-n} \\ &\leq \left(1 + |\lambda|^{-1} + \dots + |\lambda|^{1-n}\right) \max_{1 \leq i \leq n} \|B_0^{-1}B_i\| \\ &\leq \frac{1}{1 - |\lambda|^{-1}} \max_{1 \leq i \leq n} \|B_0^{-1}B_i\| \end{aligned}$$

Thus the conditions of our perturbation theorem hold and for sufficiently large $|\lambda|$, it follows $\|\lambda^{-1}(B_0^{-1}B_1 + \dots + B_0^{-1}B_n\lambda^{1-n})\| < 1$. Hence

$$I + \lambda^{-1}(B_0^{-1}B_1 + \dots + B_0^{-1}B_n\lambda^{1-n})$$

is invertible and therefore $B(\lambda)$ is also invertible. $\square$

Theorem 3.4.1. *Let $A(\lambda)$ and $B(\lambda)$ be matrix polynomials in $M_n(\mathbb{C})$ or $(M_n(R))$. Then both left and right division of $A(\lambda)$ by $B(\lambda)$ is possible and the respective quotients and remainders are unique.*

Proof. We proceed by induction on $\deg B$, and clearly if $\deg B = 0$ the result holds. If $\deg B = p > \deg A(\lambda) = m$ then the result follows simply. For, take $Q_r(\lambda) = 0$ and $R_r(\lambda) = A(\lambda)$. The conditions of right division are met. Now suppose that $p \leq m$. It is easy to see that

$$\begin{aligned} A(\lambda) &= A_0\lambda^m + A_1\lambda^{m-1} + \cdots + A_{m-1}\lambda + A_m \\ &= A_0B_0^{-1}\lambda^{m-p}B(\lambda) - \sum_{j=1}^p A_0B_0^{-1}B_j\lambda^{m-j} + \sum_{j=1}^m A_j\lambda^{m-j} \\ &= Q_1(\lambda)B(\lambda) + A_1(\lambda) \end{aligned}$$

where $\deg A_1(\lambda) < \deg A(\lambda)$. Our inductive hypothesis assumed the division was possible for matrix polynomials $A(\lambda)$ of degree $< p$. Therefore, $A_1(\lambda) = Q_2(\lambda)B(\lambda) + R(\lambda)$, where the degree of $B(\lambda) < p$. Finally, with $Q(\lambda) = Q_1(\lambda) + Q_2(\lambda)$, there results $A(\lambda) = Q(\lambda)B(\lambda) + R(\lambda)$.

To establish uniqueness we assume two right divisors and quotients have been determined. Thus

$$\begin{aligned} A &= Q_{r_1}(\lambda)B(\lambda) + R_{r_1}(\lambda) \\ A &= Q_{r_2}(\lambda)B(\lambda) + R_{r_2}(\lambda) \end{aligned}$$

Subtract to get

$$0 = (Q_{r_1}(\lambda) - Q_{r_2}(\lambda))B(\lambda) + R_{r_1}(\lambda) - R_{r_2}(\lambda).$$

If $Q_{r_1}(\lambda) - Q_{r_2}(\lambda) \neq 0$, we know the degree of $(Q_{r_1}(\lambda) - Q_{r_2}(\lambda))B(\lambda)$ is greater than the degree of $R_{r_1}(\lambda) - R_{r_2}(\lambda)$. This contradiction implies that $Q_{r_1}(\lambda) - Q_{r_2}(\lambda) = 0$, which in turn implies that $R_{r_1}(\lambda) - R_{r_2}(\lambda) = 0$. Hence the decomposition is unique. $\square$

Hamilton-Cayley Theorem

Let

$$B(\lambda) = B_0\lambda^m + B_1\lambda^{m-1} + \cdots + B_{m-1}\lambda + B_m$$

with $B_0 \neq 0$. We can also, write $B(\lambda) = \sum_{i=0}^m \lambda^{m-i} B_i$. Both versions are the same. However, when $A \in M_n(F)$, there are two possible evaluations of $B(A)$.

Definition 3.4.2. Let $B(\lambda), A \in M_n(\mathbb{C})$ (or $M_n(R)$) and $B(\lambda) = B_0\lambda^m + B_1\lambda^{m-1} + \cdots + B_{m-1}\lambda + B_m$. Define

$$\begin{aligned} B(A) &= B_0A^m + B_1A^{m-1} + \cdots + B_m && \text{“right value”} \\ \widehat{B}(A) &= A^mB_0 + A^{m-1}B_1 + \cdots + B_m && \text{“left value”} \end{aligned}$$

The generalized Bézout theorem gives the remainder of $B(\lambda)$ divided by $\lambda I - A$. In fact, we have

Theorem 3.4.2 (Generalized Bézout Theorem). *The right division of $B(\lambda)$ by $\lambda I - A$ has remainder*

$$R_r(\lambda) = B(A)$$

Similarly, the left division of $B(\lambda)$ by $(\lambda I - A)$ has remainder

$$R_\ell(\lambda) = \widehat{B}(A).$$

Proof. In the case $\deg B(\lambda) = 1$, we have

$$B_0\lambda + B_1 = B_0(\lambda I - A) + B_0A + B_1.$$

The remainder $R_r(\lambda) = B_0A + B_1 = B(A)$. Assume the result holds for all polynomials up to degree $p - 1$. We have

$$\begin{aligned} B(\lambda) &= B_0\lambda^p + B_1\lambda^{p-1} + \cdots + B_p \\ &= B_0\lambda^{p-1}(\lambda I - A) + B_0A\lambda^{p-1} + B_1\lambda^{p-1} + \cdots \\ &= B_0\lambda^{p-1}(\lambda I - A) + B_1(\lambda) \end{aligned}$$

where $\deg B_1(\lambda) \leq p - 1$. By induction

$$B(\lambda) = B_0\lambda^{p-1}(\lambda I - A) + Q_r(\lambda)(\lambda I - A) + B_1(A)$$

$B_1(A) = (B_0A + B_1)A^{p-1} + B_2A^{p-2} + \cdots + B_{p-1}A + B_p = B(A)$. This proves the result. $\square$

Corollary 3.4.1. $(\lambda I - A)$ divides $B(\lambda)$ if and only if $B(A) = 0$ (resp $\widehat{B}(A) = 0$).

Combining the Bézout result and the adjoint formulation of the matrix inverse, we can establish the important Hamilton-Cayley theorem.

Theorem 3.4.3 (Hamilton-Cayley). *Let $A \in M_n(\mathbb{C})$ (or $M_n(R)$) with characteristic polynomial $p_A(\lambda)$. Then $p_A(A) = 0$.*

Proof. Recall the adjoint formulation of the inverse as $\hat{C} = \frac{1}{\det C} C_{ij}^T = C^{-1}$. Now let $B = \text{adj}(A - \lambda I)$. Then

$$\begin{aligned} B(\lambda I - A) &= \det(\lambda I - A)I \\ (\lambda I - A)B &= \det(\lambda I - A)I. \end{aligned}$$

These equations show that $p(\lambda)I = \det(\lambda I - A)I$ is divisible on the right and the left by $(\lambda I - A)$ without remainder. It follows from the generalized Bézout theorem that this is possible only if $p_A(A) = 0$. $\square$

Let $A \in M_n(C)$. Now that we know any A satisfies its characteristic polynomial, we might also ask if there are polynomials of lower degree that it also satisfies. In particular, we will study the so-called *minimal* polynomial that a matrix satisfies. The nature of this polynomial will shed considerable light on the fundamental structure of A . For example, both matrices below have the same characteristic polynomial $P(\lambda) = (\lambda - 2)^3$.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

Hence $\lambda = 2$ is an eigenvalue of multiplicity three. However, A satisfies the much simpler first degree polynomial $(\lambda - 2)$ while there is no polynomial of degree less than three that B satisfies. By this time you recognize that A has three linearly independent eigenvectors, while B has only one eigenvector. We will take this subject up in a later chapter.

Biographies

Arthur Cayley (1821-1895), one of the most prolific mathematicians of his era and of all time, born in Richmond, Surrey, and studied mathematics at Cambridge. For four years he taught at Cambridge having won a Fellowship and, during this period, he published 28 papers in the Cambridge Mathematical Journal. A Cambridge fellowship had a limited tenure so Cayley had to find a profession. He chose law and was admitted to the bar in 1849. He spent 14 years as a lawyer, but Cayley always considered it as a means to make money so that he could pursue mathematics. During these 14 years as a lawyer Cayley published about 250 mathematical papers! Part of that time he worked in collaboration with **James Joseph Sylvester**³ (1814-

³In 1841 he went to the United States to become professor at the University of Virginia, but just four years later resigned and returned to England. He took to teaching private

1897), another lawyer. Together, but not in collaboration, they founded the algebraic theory of invariants 1843.

In 1863 Cayley was appointed Sadleirian professor of Pure Mathematics at Cambridge. This involved a very large decrease in income. However Cayley was very happy to devote himself entirely to mathematics. He published over 900 papers and notes covering nearly every aspect of modern mathematics.

The most important of his work is in developing the algebra of matrices, work in non-Euclidean geometry and n -dimensional geometry. Importantly, he also clarified many of the theorems of algebraic geometry that had previously been only hinted at, and he was among the first to realize how many different areas of mathematics were linked together by group theory.

As early as 1849 Cayley wrote a paper linking his ideas on permutations with Cauchy's. In 1854 Cayley wrote two papers which are remarkable for the insight they have of abstract groups. At that time the only known groups were groups of permutations and even this was a radically new area, yet Cayley defines an abstract group and gives a table to display the group multiplication.

Cayley developed the theory of algebraic invariance, and his development of n -dimensional geometry has been applied in physics to the study of the space-time continuum. His work on matrices served as a foundation for quantum mechanics, which was developed by Werner Heisenberg in 1925. Cayley also suggested that Euclidean and non-Euclidean geometry are special types of geometry. He united projective geometry and metrical geometry which is dependent on sizes of angles and lengths of lines.

Heinrich Weber (1842-1913) was born and educated in Heidelberg, where he became professor 1869. He then taught at a number of institutions in Germany and Switzerland. His main work was in algebra and number theory. He is best known for his outstanding text *Lehrbuch der Algebra* published in 1895.

Weber worked hard to connect the various theories even fundamental concepts such as a field and a group, which were seen as tools and not properly developed as theories in his *Die partiellen Differentialgleichungen der mathematischen Physik* 1900-01, which was essentially a reworking of a book of the same title based on lectures given by Bernhard Riemann and

pupils and had among them Florence Nightingale. By 1850 he became a barrister, and by 1855 returned to an academic life at the Royal Military Academy in Woolwich, London. He returned to the US again in 1877 to become professor at the new Johns Hopkins University, but returned to England once again in 1877. Sylvester coined the term 'matrix' in 1850.

written by Karl Hattendorff.

Etienne Bézout (1730-1783) was a mathematician who represents a characteristic aspect of the subject at that time. One of the many successful textbook projects produced in the 18th century was Bézout's *Cours de mathématique*, a six volume work that first appeared in 1764-1769, which was almost immediately issued in a new edition of 1770-1772, and which boasted many versions in French and other languages. (The first American textbook in analytic geometry, incidentally, was derived in 1826 from Bézout's *Cours*.)

It was through such compilations, rather than through the original works of the authors themselves, that the mathematical advances of Euler and d'Alembert became widely known. Bézout's name is familiar today in connection with the use of determinants in algebraic elimination. In a memoir of the Paris Academy for 1764, and more extensively in a treatise of 1779 entitled *Theorie generale des equations algebriques*, Bézout gave artificial rules, similar to Cramer's, for solving n simultaneous linear equations in n unknowns. He is best known for an extension of these to a system of equations in one or more unknowns in which it is required to find the condition on the coefficients necessary for the equations to have a common solution. To take a very simple case, one might ask for the condition that the equations $a_1x + b_1y + c_1 = 0$, $a_2x + b_2y + c_2 = 0$, $a_3x + b_3y + c_3 = 0$ have a common solution. The necessary condition is that the eliminant a special case of the "Bezoutiant," should be 0.

Somewhat more complicated eliminants arise when conditions are sought for two polynomial equations of unequal degree to have a common solution. Bézout also was the first one to give a satisfactory proof of the theorem, known to Maclaurin and Cramer, that two algebraic curves of degrees m and n respectively intersect in general in $m \cdot n$ points; hence, this is often called Bézout's theorem. Euler also had contributed to the theory of elimination, but less extensively than did Bézout.

Taken from *A History of Mathematics* by Carl Boyer

William Rowen Hamilton**Born Aug. 3/4, 1805, Dublin, Ire. and died Sept. 2, 1865, Dublin**

Irish mathematician and astronomer who developed the theory of quaternions, a landmark in the development of algebra, and discovered the phenomenon of conical refraction. His unification of dynamics and optics, moreover, has had lasting influence on mathematical physics, even though the full significance of his work was not fully appreciated until after the rise of quantum mechanics.

Like his English contemporaries Thomas Babington Macaulay and John Stuart Mill, **Hamilton** showed unusual intellect as a child. Before the age of three his parents sent him to live with his father's brother, James, a learned clergyman and schoolmaster at an Anglican school at Trim, a small town near Dublin, where he remained until 1823, when he entered Trinity College, Dublin. Within a few months of his arrival at his uncle's he could read English easily and was advanced in arithmetic; at five he could translate Latin, Greek, and Hebrew and recite Homer, Milton and Dryden. Before his 12th birthday he had compiled a grammar of Syriac, and by the age of 14 he had sufficient mastery of the Persian language to compose a welcome to the Persian ambassador on his visit to Dublin.

Hamilton became interested in mathematics after a meeting in 1820 with Zerah Colburn, an American who could calculate mentally with astonishing speed. Having read the *Eléments d'algèbre* of Alexis-Claude Clairaut and Isaac Newton's *Principia*, **Hamilton** had immersed himself in the five volumes of Pierre-Simon Laplace's *Traité de mécanique céleste* (1798-1827; *Celestial Mechanics*) by the time he was 16. His detection of a flaw in Laplace's reasoning brought him to the attention of John Brinkley, professor of astronomy at Trinity College. When **Hamilton** was 17, he sent Brinkley, then president of the Royal Irish Academy, an original memoir about geometrical optics. Brinkley, in forwarding the memoir to the Academy, is said to have remarked: "This young man, I do not say *will be*, but *is*, the first mathematician of his age."

In 1823 **Hamilton** entered Trinity College, from which he obtained the highest honours in both classics and mathematics. Meanwhile, he continued his research in optics and in April 1827 submitted this "theory of Systems of Rays" to the Academy. The paper transformed geometrical optics into a new mathematical science by establishing one uniform method for the solution of all problems in that field. **Hamilton** started from the principle, originated by the 17th-century French mathematician Pierre de Fermat, that light takes the shortest possible time in going from one point to another, whether the path is straight or is bent by refraction. **Hamilton's** key idea

was to consider the time (or a related quantity called the “action”) as a function of the end points between which the light passes and to show that this quantity varied when the coordinates of the end points varied, according to a law that he called the law of varying action. He showed that the entire theory of systems of rays is reducible to the study of this characteristic function.

Shortly after **Hamilton** submitted his paper and while still an undergraduate, Trinity College elected him to the post of Andrews professor of astronomy and royal astronomer of Ireland, to succeed Brinkley, who had been made bishop. Thus an undergraduate (not quite 22 years old) became ex officio an examiner of graduates who were candidates for the Bishop Law Prize in mathematics. The electors’ object was to provide **Hamilton** with a research post free from heavy teaching duties. Accordingly, in October 1827 **Hamilton** took up residence next to Dunsink Observatory, 5 miles (8 km) from Dublin, where he lived for the rest of his life. He proved to be an unsuccessful observer, but large audiences were attracted by the distinctly literary flavour of his lectures on astronomy. Throughout his life **Hamilton** was attracted to literature and considered the poet **William** Wordsworth among his fiends, although Wordsworth advised him to write mathematics rather than poetry.

With eigenvalues we are able to begin spectral analysis. That part is the derivation of the various normal forms for matrices. We begin with a relatively weak form of the Jordan form, which is coming up.

First of all, as you have seen diagonal matrices furnish the easiest form for matrix analysis. Also, linear systems are very simple to solve for diagonal matrices. The next simplest class of matrices are the triangular matrices. We begin with the following result based on the idea of similarity.

3.5 Similarity

Definition 3.5.1. A matrix $B \in M_n$ is said to be *similar* to $A \in M_n$ if there exists a nonsingular matrix $S \in M_n$ such that

$$B = S^{-1}AS.$$

The transformation $A \rightarrow S^{-1}AS$ is called a *similarity transformation*. Sometimes we write $A \sim B$. Note that similarity is an *equivalence* relation:

- | | | |
|-------|--|--------------|
| (i) | $A \sim A$ | reflexivity |
| (ii) | $B \sim A \Rightarrow A \sim B$ | symmetry |
| (iii) | $B \sim A$ and $A \sim C \Rightarrow B \sim C$ | transitivity |

Theorem 3.5.1. *Similar matrices have the same characteristic polynomial.*

Proof. We suppose A, B , and $S \in M_n$ with S invertible and $B = S^{-1}AS$. Then

$$\begin{aligned}\lambda I - B &= \lambda I - S^{-1}AS \\ &= \lambda S^{-1}IS - S^{-1}AS \\ &= S^{-1}(\lambda I - A)S.\end{aligned}$$

Hence

$$\begin{aligned}\det(\lambda I - B) &= \det(S^{-1}(\lambda I - A)S) \\ &= \det(S^{-1}) \det(\lambda I - A) \det S \\ &= \det(\lambda I - A)\end{aligned}$$

because $1 = \det I = \det(SS^{-1}) = \det S \det S^{-1}$. □

A simple consequence of Theorem 3.5.1 and Corollary 2.3.1 follows.

Corollary 3.5.1. *If A and B are in M_n and if A and B are similar, then they have the same eigenvalues counted according to multiplicity, and therefore the same rank.*

We remark that even though $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ have the same eigenvalues, 0 and 0, they are not similar. Hence the converse is not true.

Another immediate corollary of Theorem 3.5.1 can be expressed in terms of the invariance of the trace and determinant of *similarity transformations*, that is functions on $M_n(F)$ defined for any invertible matrix S by $T_S(A) = S^{-1}AS$. Note that such transformations are linear mappings from $M_n(F) \rightarrow M_n(F)$. We shall see how important are those properties of matrices that are invariant (i.e. do not change) under similarity transformations.

Corollary 3.5.2. *If $A, B \in M_n$ are similar, then they have the same trace and same determinant. That is, $\text{tr}(A) = \text{tr}(B)$ and $\det A = \det B$.*

Theorem 3.5.2. *If $A \in M_n$, then A is similar to a triangular matrix.*

Proof. The following sketch shows the first two steps of the proof. A formal induction can be applied to achieve the full result.

Let λ_1 be an eigenvalue of A with eigenvector u_1 . Select a basis, say S_1 , of $\mathbb{C}_n$ and arrange these vectors into columns of the matrix P_1 , with u_1

in the first column. Define $B_1 = P_1^{-1}AP_1$. Then B_1 is the representation of A in the new basis and so

$$B_1 = \begin{bmatrix} \lambda & \alpha_1 & \dots & \alpha_{n-1} \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix} \quad \text{where } A_2 \text{ is } (n-1) \times (n-1)$$

because $Au_1 = \lambda_1 u_1$. Remembering that B_1 is the representation of A in the basis S_1 , then $[u_1]_{S_1} = e_1$ and hence $B_1 e_1 = \lambda e_1 = \lambda_1 [1, 0, \dots, 0]^T$. By similarity, the characteristic polynomial of B_1 is the same as that of A , but more importantly (using expansion by minors down the first column)

$$\det(\lambda I - B_1) = (\lambda - \lambda_1) \det(\lambda I_{n-1} - A_2).$$

Now select an eigenvalue λ_2 of A_2 and pertaining eigenvector $v_2 \in \mathbb{C}_n$; so $A_2 v_2 = \lambda_2 v_2$. Note that λ_2 is also an eigenvalue of A . With this vector we define

$$\hat{u}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \hat{u}_2 = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}.$$

Select a basis of $\mathbb{C}_{n-1}$, with v_2 selected first and create the matrix $\hat{P}_2$ with this basis as columns

$$P_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \hat{P}_2 & \\ 0 & & & \end{bmatrix}.$$

It is an easy matter to see that P_2 is invertible and

$$\begin{aligned} B_2 &= P_2^{-1} B_1 P_2 \\ &= \begin{bmatrix} \lambda_1 & * & \dots & \dots & * \\ 0 & \lambda_2 & * & \dots & * \\ 0 & 0 & & A_3 & \\ \vdots & \vdots & & & \\ 0 & 0 & & & \end{bmatrix}. \end{aligned}$$

Of course, $B_2 \sim B_1$, and by the transitivity of similarity $B_2 \sim A$. Continue this process, deriving ultimately the triangular matrix $B_n \sim B_{n-1}$ and hence $B_n \sim A$. This completes the proof. $\square$

Definition 3.5.2. We say that $A \in M_n$ is *diagonalizable* if A is similar to a diagonal matrix.

Suppose $P \in M_n$ is nonsingular and $D \in M_n$ is a diagonal matrix. Let $A = PDP^{-1}$. Suppose the columns of P are the vectors $v_1, v_2, \dots, v_n$. Then

$$\begin{aligned} Av_j &= PDP^{-1}v_j \\ &= PDe_j = \lambda_j Pe_j = \lambda_j v_j \end{aligned}$$

where λ_j is the j^{th} diagonal element of D , and e_j is the j^{th} standard vector.

Similarly, if $u_1, \dots, u_n$ are n linearly independent vectors of A pertaining to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, then with Q , the matrix with columns $u_1 \dots u_n$, we have

$$AQ = QD$$

where

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Therefore

$$Q^{-1}AQ = D.$$

We have thus established the

Theorem 3.5.3. *Let $A \in M_n$. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.*

As a practical measure, the conditions of this theorem are remarkably difficult to verify.

Example 3.5.1. The matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is *not* diagonalizable.

Solution. First we note that the spectrum $\sigma(A) = \{0\}$. Solving

$$Ax = 0$$

we see that $x = c(1, 0)^T$ is the only solution. That is to say, there are *not* two linearly independent eigenvectors. hence the result.

Corollary 3.5.3. *If $A \in M_n$ is diagonalizable and B is similar to A , then B is diagonalizable.*

Proof. The proof follows directly from transitivity of similarity. However, more directly, suppose that $B \sim A$ and S is the invertible matrix such that $B = SAS^{-1}$. Then $BS = SA$. If u is an eigenvector of A with eigenvalue λ , then $SAu = \lambda Su$ and therefore $BSu = \lambda Su$. This is valid for each eigenvector. We see that if $u_1, \dots, u_n$ are the eigenvectors of A , then $Su_1, \dots, Su_n$ are the eigenvectors of B . The similarity matrix converts the eigenvectors of one matrix to the eigenvectors of the transformed matrix. $\square$

In light of these remarks, we see that similarity transforms preserve completely the dimension of eigenspaces. It is just as significant to note that if a similarity transformation diagonalizes a given matrix A , the similarity matrix must consist of the eigenvectors of A .

Corollary 3.5.4. *Let $A \in M_n$ be nonzero and nilpotent. Then A is not diagonalizable.*

Proof. Suppose $A \sim T$, where T is triangular. Since A is nilpotent (i.e. $A^m = 0$), T is nilpotent, as well. Therefore the diagonal entries of T are zero. The spectrum of T and hence A is zero, its null space has dimension n . Therefore, by Theorem 2.3.3(f), its rank is zero. Therefore $T = 0$, and hence $A = 0$. The result is proved. $\square$

Alternatively, if the nilpotent matrix A is similar to a diagonal matrix with zero diagonal entries, then A is similar to the zero matrix. Thus A is itself the zero matrix. From the obvious fact that the power of a similarity transformation is the similarity transformation of the power of a matrix, that is $(SAS^{-1})^m = SA^mS^{-1}$ (see Exercise 26), we have

Corollary 3.5.5. *Suppose $A \in M_n(\mathbb{C})$ is diagonalizable. (i) Then A^m is diagonalizable for every positive integer m . (ii) If $p(\cdot)$ is any polynomial, then $p(A)$ is diagonalizable.*

Corollary 3.5.6. *If all the eigenvalues of $A \in M_n(\mathbb{C})$ are distinct, then A is diagonalizable.*

Proposition 3.5.1. *Let $A \in M_n(\mathbb{C})$ and $\varepsilon > 0$. Then for any matrix norm $\|\cdot\|$ there is a matrix B with norm $\|B\| < \varepsilon$ for which $A+B$ is diagonalizable and for each $\lambda \in \sigma(A)$ there is a $\mu \in \sigma(B)$ for which $|\lambda - \mu| < \varepsilon$.*

Proof. First triangularize A to T by a similarity transformation, $T = SAS^{-1}$. Now add to T any diagonal matrix D so that the resulting triangular matrix $T + D$ has all distinct values. Moreover this can be accomplished by a diagonal matrix of arbitrarily small norm for any matrix norm. Then we have

$$\begin{aligned} S^{-1}(T + D)S &= S^{-1}TS + S^{-1}DS \\ &= A + B \end{aligned}$$

where $B := S^{-1}DS$. Now by the submultiplicative property of matrix norms $\|B\| \leq \|S^{-1}\| \|D\| \|S\|$. Thus to obtain the estimate $\|B\| < \varepsilon$, it is sufficient to take $\|D\| \leq \frac{\varepsilon}{\|S^{-1}\| \|S\|}$, which is possible as established above. Since the spectrum $\sigma(A + B)$ of $A + B$ has n distinct values, it follows that $A + B$ is diagonalizable. $\square$

There are many, many results on diagonalizable and non-diagonalizable matrices. Here is an interesting class of nondiagonalizable matrices we will encounter later.

Proposition 3.5.2. *pro Every matrix of the form*

$$A = \lambda I + N$$

where N is nilpotent and is not zero is not diagonalizable.

An important subclass has the form:

$$A = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \ddots & 1 \\ & \bigcirc & & & \lambda \end{bmatrix} \quad \text{“Jordan block”}.$$

Eigenvectors

Once an eigenvalue is determined, it is a relatively simple matter to find the pertaining eigenvectors. Just solve the homogeneous system $(\lambda I - A)x = 0$. Eigenvectors have a more complex structure and their study merits our attention.

Facts

- (1) $\sigma(A) = \sigma(A^T)$ including multiplicities.

(2) $\overline{\sigma(A)} = \sigma(A^*)$, including multiplicities.

Proof. $\det(\lambda I - A) = \det((\lambda I - A)^T) = \det(\lambda I - A^T)$. Similarly for A^* . $\square$

Definition 3.5.3. The linear space spanned by all the eigenvectors pertaining to an eigenvalue λ is called the *eigenspace* corresponding to the eigenvalue λ .

For any $A \in M_n$ any subspace $V \in \mathbb{C}_n$ for which

$$AV \subset V$$

is called an *invariant subspace* of A . The determination of invariant subspaces for linear transformations has been an important question for decades.

Example 3.5.2. For any upper triangular matrix T the spaces $V_j = \mathfrak{S}(e_1, e_2, \dots, e_j)$, $j = 1, \dots, n$ are invariant.

Example 3.5.3. Given $A \in M_n$, with eigenvalue λ . The eigenspace corresponding to λ and all of its subspaces are invariant subspaces. Corollary to this, the null space $N(A) = \{x | Ax = 0\}$ is an invariant subspace of A corresponding to the eigenvalue $\lambda = 0$.

Definition 3.5.4. Let $A \in M_n$ with eigenvalue λ . The dimension of the eigenspace corresponding to λ is called the *geometric multiplicity* of λ . The multiplicity of λ as a zero of the characteristic polynomial $p_A(\lambda)$ is called the *algebraic multiplicity*.

Theorem 3.5.4. If $A \in M_n$ and λ_0 is an eigenvalue of A with geometric and algebraic multiplicities m_g and m_a , respectively. Then

$$m_g \leq m_a.$$

Proof. Let $u_1 \dots u_{m_g}$ be linearly independent eigenvectors pertaining to λ_0 . Let S be a matrix consisting of a basis of $\mathbb{C}_n$ with $u_1 \dots u_{m_g}$ selected among them and placed in the first m_g columns. Then

$$B = S^{-1}AS = \begin{bmatrix} \lambda_0 I & \vdots & * \\ \dots & \dots & \dots \\ 0 & \vdots & \widehat{B} \end{bmatrix}$$

where $I = I_{m_g}$. It is easy to see that

$$p_B(\lambda) = p_A(\lambda) = (\lambda - \lambda_0)^{m_g} p_{\widehat{B}}(\lambda),$$

whence the algebraic multiplicity $m_a \geq m_g$. $\square$

Example 3.5.4. For $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, the algebraic multiplicity of a is 2, while the geometric multiplicity of 0 is 1. Hence, the equality $m_g = m_a$ is not possible, in general.

Theorem 3.5.5. *If $A \in M_n$ and for each eigenvalue $\mu \in \sigma(A)$, $m_g(\mu) = m_a(\mu)$, then A is diagonalizable. The converse is also true.*

Proof. Extend the argument given in the theorem just above. Alternatively, we can see that A must have n linearly independent eigenvalues, which follows from the $\square$

Lemma 3.5.1. *If $A \in M_n$ and $\mu \neq \lambda$ are eigenvalues with eigenspaces E_μ and E_λ respectively. Then E_μ and E_λ are linearly independent.*

Proof. Suppose $u \in E_\mu$ can be expressed as

$$u = \sum_{j=1}^k c_j v_j$$

where, of course $u \neq 0$ and $v_j \neq 0, j = 1, \dots, k$ where $v_1 \dots v_k \in E_\lambda$. Then

$$\begin{aligned} Au &= A \sum c_j v_j \\ \Rightarrow \mu u &= \lambda \sum c_j v_j \\ \Rightarrow u &= \frac{\lambda}{\mu} \sum c_j v_j, \text{ if } \mu \neq 0. \end{aligned}$$

Since $\lambda/\mu \neq 1$, we have a contradiction. If $\mu = 0$, then $\sum c_j v_j = 0$ and this implies $u = 0$. In either case we have a contradiction, the result is therefore proved.

While both A and A^T have the same eigenvalues, counted even with multiplicities, the eigenspaces can be very much different. Consider the example where

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad A^T = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}.$$

The eigenvalue $\lambda = 1$ has eigenvector $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for A and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for A^T . $\square$

A new concept of left eigenvector yield some interesting results.

Definition 3.5.5. We say λ is a *left eigenvector* of $A \in M_n$ if there is a vector $y \in \mathbb{C}_n$ for which

$$y^* A = \lambda y^*$$

or in $\mathbb{R}_n$ if $y^T A = \lambda y^T$. Taking adjoints the two sides of this equality become

$$\begin{aligned}(y^* A)^* &= A^* y \\ (\lambda y^*)^* &= \bar{\lambda} y.\end{aligned}$$

Putting these lines together $A^* y = \bar{\lambda} y$ and hence $\bar{\lambda}$ is an eigenvalue of A^* . Here's the big result.

Theorem 3.5.6. *Let $A \in M_n(\mathbb{C})$ with eigenvalues λ and eigenvector u . Let u, v be left eigenvectors pertaining to μ, λ , respectively. If $\mu \neq \lambda$, $v^* u = \langle u, v \rangle = 0$.*

Proof. We have

$$v^* u = \frac{1}{\mu} v^* A u = \frac{\lambda}{\mu} v^* u.$$

Assuming $\mu \neq 0$ we have a contradiction. If $\mu = 0$, argue as $v^* u = \frac{1}{\lambda} v^* A u = \frac{\mu}{\lambda} v^* u = 0$, which was to be proved. The result is proved. $\square$

Note that left eigenvectors of A are (right) eigenvectors of A^* (A^T in the real case).

3.6 Equivalent norms and convergent matrices

Equivalent norms

So far we have defined a number of different norms. Just what “different” means is subject to different interpretations. For example, we might agree that different means that the two norms have a different value for some matrix A . On the other hand if we have two matrix norms $\|\cdot\|_a$ and $\|\cdot\|_b$, we might be prepared to say that if for two positive constants $m, M > 0$

$$0 < m \leq \frac{\|A\|_a}{\|A\|_b} \leq M \quad \text{for all } A \in M_n(F)$$

then these norms are not really so different but rather are equivalent because “small” in one norm implies “small” in the other and the same for “large.” Indeed, when two norms satisfy the condition above we will call them equivalent. The remarkable fact is that all subordinate norms on a finite dimensional space are equivalent. This is a direct consequence of the similar result for vector norms. We state it below but leave the details of the proof to the reader.

Theorem 3.6.1. *Any two matrix norms, $\|\cdot\|_a$ and $\|\cdot\|_b$, on $M_n(\mathbb{C})$ are equivalent in the sense that there exist two positive constants $m, M > 0$*

$$0 < m \leq \frac{\|A\|_a}{\|A\|_b} \leq M \quad \text{for all } A \in M_n(F)$$

We have already proved one convergence result about the invertibility of $I - A$ when $\|A\| < 1$. A deeper version of this result can be proved based on a new matrix norm. This result has important consequences in general matrix theory and particularly in computational matrix theory. It is most important in applications, where having $\rho(A) < 1$ can yield the same results as having $\|A\| < 1$.

Lemma 3.6.1. *Let $\|\cdot\|$ be a matrix norm that is subordinate to the vector norm (on C_n) $\|\cdot\|$. Then for each $A \in M_n$*

$$\rho(A) \leq \|A\|.$$

Note: We use the same notation for both vector and matrix norms.

Proof. Let $\lambda \in \sigma(A)$. Then with corresponding eigenvector x_λ we have $Ax_\lambda = \lambda x_\lambda$. Normalizing x_λ so that $\|x_\lambda\| = 1$, we have $\|A\| = \max_{\|x\|=1} \|Ax\| \geq \|Ax_\lambda\| = |\lambda| \|x_\lambda\| = |\lambda|$. Hence

$$\|A\| \geq \max_{\lambda \in \sigma(A)} |\lambda| = \rho(A).$$

□

Theorem 3.6.2. *For each $A \in M_n$ and $\varepsilon > 0$ there is a vector norm $\|\cdot\|$ on C_n for which the subordinate matrix norm $\|\cdot\|$ satisfies*

$$\rho(A) \leq \|A\| \leq \rho(A) + \varepsilon.$$

Proof. The proof follows in a series of simple steps, the first of which is interesting in its own right. First we know that A is similar to a triangular matrix B —from a previous theorem

$$B = SAS^{-1} = \Lambda + U$$

where Λ is the diagonal part of B and U is strictly upper triangular part of B , with zeros filled in elsewhere.

Note that the diagonal of B is the spectrum of B and hence that of A . This is important. Now select a $\delta > 0$ and form the diagonal matrix

$$D = \begin{bmatrix} 1 & & & \\ & \delta^{-1} & & \circ \\ & & \ddots & \\ & \circ & & \delta^{1-n} \end{bmatrix}.$$

A brief computation reveals that

$$\begin{aligned} C &= DBD^{-1} = DSAS^{-1}D^{-1} \\ &= D(\Lambda + U)D^{-1} \\ &= \Lambda + DUD^{-1} &= \Lambda + V \end{aligned}$$

where $V = DUD^{-1}$ and more specifically $v_{ij} = \delta^{j-i}u_{ij}$ for $j > i$ and of course Λ is a diagonal matrix with the spectrum of A for the diagonal elements. In this way we see that for δ small enough we have arranged that A is similar to the diagonal matrix of its spectral elements plus a small triangular perturbation. We now define the new vector norm on $\mathbb{C}_n$ by

$$\|x\|_A = \langle (DS)^*(DS)x, x \rangle^{1/2}.$$

We recognise this to be a norm from a previous result. Now compute the matrix norm

$$\|A\|_A = \max_{\|x\|_A=1} \|Ax\|_A.$$

Thus

$$\begin{aligned} \|Ax\|_A^2 &= \langle DSAx, DSAx \rangle \\ &= \langle CDSx, CDSx \rangle \\ &\leq \|C\|_2^2 \|DSx\|^2 \\ &= \|C^*C\|_2 \|x\|_A^2. \end{aligned}$$

From $C = \Lambda + V$, it follows that

$$C^*C = \Lambda^*\Lambda + \Lambda^*V + V^*\Lambda + V^*V.$$

Because the last three terms have a δ^p multiplier for various **positive** powers p , we can make the terms $\Lambda^*V + V^*\Lambda + V^*V$ small in *any* norm by taking δ sufficiently small. Also the diagonal elements of $\Lambda^*\Lambda$ have the form

$$|\lambda_i|^2 \quad \lambda_i \in \mathfrak{S}(A).$$

Talking δ sufficiently small so that each of $\|\Lambda^*V\|_2$, $\|V^*\Lambda\|_2$ and $\|V^*V\|_2$ is less than $\epsilon/3$. With $\|x\|_A^2 = 1$ as required, we have

$$\|Ax\|_A \leq (\rho(A) + \varepsilon)\|x\|_A$$

and we know already that $\|A\|_A \geq \rho(A)$. $\square$

An important corollary places a lower bound on vector norms is given below.

Corollary 3.6.1. *For any $A \in M_n$*

$$\rho(A) = \inf_{\|\cdot\|} (\max_{\|x\|=1} \|Ax\|)$$

where the infimum is taken over all vector norms.

Convergent matrices

Definition 3.6.1. We say that a matrix $A \in M_n(F)$ for $F = \mathbb{R}$ or $\mathbb{C}$ is *convergent* if

$$\lim_{m \rightarrow \infty} A^m = 0 \quad \leftarrow \text{the zero matrix.}$$

That is, for each $1 \leq i, j \leq n$ the $\lim_{m \rightarrow \infty} (A^m)_{ij} = 0$. This is sometimes called *pointwise convergence*.

Theorem 3.6.3. *The following three statements are equivalent:*

- (a) A is convergent.
- (b) $\lim_{n \rightarrow \infty} \|A^n\| = 0$, for some matrix norm.
- (c) $\rho(A) < 1$.

Proof. These results follow substantially from previously proven results. However, for completeness, assume (a) holds. Then

$$\lim_{n \rightarrow \infty} \max_{ij} |(A^n)_{ij}| = 0$$

or what is the same we have

$$\lim_{n \rightarrow \infty} \|A^n\|_\infty = 0$$

which is (b). Now suppose that (b) holds. If $\rho(A) \geq 1$ there is a vector x for which $\|A^n x\| = \|\rho(A)\|^n \|x\|$. Therefore $\|A^n\| \geq 1$, which contradicts (b). Thus (c) holds. Next if (c) holds we can apply the above theorem to establish that there is a norm for which $\|A\| < 1$. Hence (b) follows. Finally, we know that

$$\|A^m\|_\infty \leq M\|A^m\| \quad (\star)$$

whence $\lim_{n \rightarrow \infty} A^n = 0$. (That is (a) $\equiv$ (b)). In sum we have shown that (a) $\equiv$ (b) and (b) $\equiv$ (c). $\square$

To establish $(\star)$ we need the following.

Theorem 3.6.4. *If $\|\cdot\|$ and $\|\cdot\|'$ are two vector norms on $\mathbb{C}_n$ then there are constants m and M so that*

$$m\|x\|' \leq \|x\| < M\|x\|'$$

for all $x \in \mathbb{C}_n$.

This result carries over to matrix norms subordinate to vector norms by simple inheritance. To prove this result we use compactness ideas. Suppose there is a sequence of vectors x_j for which

$$1 \leq \|x_j\| < \frac{1}{j} \|x_j\|'$$

and for which the components of x_j are bounded in modulus. By compactness there is a convergent subsequence of the x_j , for which $\lim x_j = x \neq 0$. But $\|x\|' < \infty$. Hence $\|x\| = 0$, a contradiction.

Here is the new and improved version of our previous result.

Theorem 3.6.5. *If $\rho(A) < 1$ then $(I - A)^{-1}$ exists and*

$$(I - A)^{-1} = I + A + A^2 + \cdots .$$

Proof. Select a matrix norm $\|\cdot\|$ for which $\|A\| < 1$. Apply previous calculations. We know that every matrix can be triangularized and “almost” diagonalized we may ask if we can eliminate altogether the off-diagonal terms. The answer is unfortunately **no**. But we can resolve the diagonal question completely. $\square$

3.7 Exercises

1. Suppose that $U \in M_n$ is orthogonal and let $\|\cdot\|$ be a matrix norm. Show that $\|U\| \geq 1$.
2. In two dimensions the counter-clockwise rotations through the angle θ are given by

$$B_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Find the eigenvalues and eigenvectors for all θ . (Note the two special cases, θ is not equal to an even multiple of π and $\theta = 0$.)

3. Given two matrices $A, B \in M_n(\mathbb{C})$. Define the *commutant* of A and B by $[A, B] = AB - BA$. Prove that $\text{tr}[A, B] = 0$.
4. Given two finite sequences $\{c_k\}_{k=1}^n$ and $\{d_k\}_{k=1}^n$. Prove that $\sum_k |c_k d_k| \leq \max_k |d_k| \sum_k |c_k|$.
5. Verify that a matrix norm which is subordinate to a vector norm satisfies norm conditions (i) and (ii).
6. Let $A \in M_n(\mathbb{C})$. Show that the matrix norm subordinate to the vector norm $\|\cdot\|_\infty$ is given by

$$\|A\|_\infty = \max_i \|r_i(A)\|_1$$

where as usual $r_i(A)$ denotes the i^{th} row of the matrix A and $\|\cdot\|_1$ is the ℓ_1 norm.

7. The Hilbert matrix, H_n of order n is defined by

$$h_{ij} = \frac{1}{i+j-1} \quad 1 \leq i, j \leq n$$

8. Show that $\|H_n\|_1 < \ln n$.
9. Show that $\|H_n\|_\infty = 1$.
10. Show that $\|H_n\|_2 \sim n^{\frac{1}{2}}$.
11. Show that the spectral radius of H_n is bounded by 1.
12. Show that for each $\varepsilon > 0$ there exists an integer N such that if $n > N$ there is a vector $x \in R_n$ with $\|x\|_2=1$ such that $\|H_n x\|_2 < \varepsilon$.

13. Same as the previous question except that you need to show that $N = \left\lfloor \frac{1}{\varepsilon^{1/2}} \right\rfloor + 1$ will work.
14. Show that the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & -1 & 2 \end{bmatrix}$ is not diagonalizable. Let $A \in M_n(\mathbb{C})$.
15. We know that the characteristic polynomial of a matrix $A \in M_{12}(\mathbb{C})$ is equal to $p_A(\lambda) = (\lambda - 1)^{12} - 1$. Show that A is not similar to the identity matrix.
16. We know that the spectrum of $A \in M_3(\mathbb{C})$ is $\sigma(A) = \{1, 1, -2\}$ and the corresponding eigenvectors are $\{[1, 2, 1]^T, [2, 1, -1]^T, [1, 1, 2]^T\}$.
- (i) Is it possible to determine A ? Why or why not? If so, prove it. If not show two different matrices with the given spectrum and eigenvectors.
 - (ii) Is A diagonalizable?
17. Prove that if $A \in M_n(\mathbb{C})$ is diagonalizable then for each $\lambda \in \sigma(A)$, the algebraic and geometric multiplicities are equal. That is, $m_a(\lambda) = m_g(\lambda)$.
18. Show that $B^{-1}(\lambda)$ exists for $|\lambda|$ sufficiently large.
19. We say that a matrix $A \in M_n(R)$ is row stochastic if all its entries are non negative and the sum of the entries of each row is one.
- (i) Prove that $1 \in \sigma(A)$.
 - (ii) Prove that $\rho(A) = 1$.
20. We say that $A, B \in M_n(\mathbb{C})$ are *quasi-commutative* if the spectrum of $AB - BA$ is just zero, i.e. $\sigma(AB - BA) = \{0\}$.
21. Prove that if AB is nilpotent then so is BA .
22. A matrix $A \in M_n(\mathbb{C})$ has a square root if there is a matrix $B \in M_n(\mathbb{C})$ such that $B^2 = A$. Show that if A is diagonalizable then it has a square root.
23. Prove that every matrix that commutes with every diagonal matrix is itself diagonal.

24. Suppose that if $A \in M_n(\mathbb{C})$ is diagonalizable and for each $\lambda \in \sigma(A)$, $|\lambda| < 1$. Prove directly from similarity ideas that $\lim_{n \rightarrow \infty} A^n = 0$. (That is, do not apply the more general theorem from the lecture notes.)
25. We say that A is *right quasi-invertible* if there exists a matrix $B \in M_n(\mathbb{C})$ such that $AB \sim D$, where D is a diagonal matrix with diagonal entries nonzero. Similarly, we say that A is *left quasi-invertible* if there exists a matrix $B \in M_n(\mathbb{C})$ such that $BA \sim D$, where D is a diagonal matrix with diagonal entries nonzero.
- (i) Show that if A is *right quasi-invertible* then it is invertible.
- (ii) Show that if A is *right quasi-invertible* then it is *left quasi-invertible*.
- (iii) Prove that *quasi-invertibility* is not an equivalence relation. (Hint. How do we usually show that an assertion is false?)
26. Suppose that $A, S \in M_n$, with S invertible, and m is a positive integer. Show that $(SAS^{-1})^m = SA^mS^{-1}$.
27. Consider the rotation

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

Show that the eigenvalues are the same as for the rotation

$$\begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}$$

See Example 2 of section 3.2.

28. Let $A \in M_n(\mathbb{C})$. Define $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$. (i) Prove that e^A exists. (ii) Suppose that $A, B \in M_n(\mathbb{C})$. Is $e^{A+B} = e^A e^B$? If not, when is it true?
29. Suppose that $A, B \in M_n(\mathbb{C})$. We say $A \smile B$ if $[A, B] = 0$, where $[\cdot, \cdot]$ is the commutator. Prove or disprove that " $\smile$ " is an equivalence relation. Answer the same question in the case of quasi-commutativity. (See Exercise 20)
30. Let $u, v \in \mathbb{C}_n$. Find $(I + uv^*)^m$.

31. Define the function f on $M_{m \times n}(\mathbb{C})$ by $f(A) = \text{rank } A$, and suppose that $\|\cdot\|$ is any matrix norm. Prove the following. (1) If for any matrix A for which $f(A) = n$ show that f is continuous in $\|\cdot\|$. That is, for every $\varepsilon > 0$ then $f(B) = n$ for every matrix B with $\|B - A\| < \varepsilon$. f is continuous in an ε -neighborhood of A . That is, for any B with $\|A - B\| < \varepsilon$, then $f(B) = f(A)$. (This deviates slightly from the usual definition of continuity because the function f is integer valued.) (2) If $f(A) < n$, then f is not continuous in every ε -neighborhood of A .

3.8 Appendix A

It is desired to solve the equation $p(\lambda) = 0$ for coefficients in $p_0, \dots, p_n \in \mathbb{C}$ or R . The basic theorem on this subject is called the Fundamental Theorem of Algebra (FTA), whose importance is manifest by its hundreds of applications. Concomitant with the FTA is the notion of reducible and irreducible factors.

Theorem 3.8.1 (Theorem Fundamental Theorem of Algebra). *Given any polynomial $p(\lambda) = p_0\lambda^n + p_1\lambda^{n-1} + \dots + p_n$ with coefficients in $p_0, \dots, p_n \in \mathbb{C}$. There is at least one solution $\lambda \in \mathbb{C}$ to the equation $p(\lambda) = 0$.*

Though proving this result would take us too far afield of our subject, we remark that the simplest proof of this result no doubt comes as a direct application of Liouville's Theorem, a result in complex analysis. As a corollary to the FTA, we have that there are exactly n solutions to $p(\lambda) = 0$ when counted with multiplicity. Proved originally by Gauss, the proof of this theorem eluded mathematicians for many years. Let us assume that $p_0 = 1$ to make the factorization simpler to write. Thus

$$p(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{m_i} \quad (4)$$

As we know, in the case that the coefficients $p_0, \dots, p_n$ are real, there may be complex solutions. As is easy to prove, complex solutions must come in complex conjugate pairs. For if $\lambda = r + is$ is a solution

$$p(r + is) = (r + is)^n + p_1(r + is)^{n-1} + \dots + p_{n-1}(r + is) + p_n = 0$$

Because the coefficients are real, the real and imaginary parts of the powers $(r + is)^j$ remain respectively real or imaginary upon multiplication by p_j .

Thus

$$\begin{aligned} p(r + is) &= \operatorname{Re}(r + is)^n + p_1 \operatorname{Re}(r + is)^{n-1} + \cdots + p_{n-1} \operatorname{Re}(r + is) + p_n \\ &\quad + i \left(\operatorname{Im}(r + is)^n + p_1 \operatorname{Im}(r + is)^{n-1} + \cdots + p_{n-1} \operatorname{Im}(r + is) + p_n \right) \\ &= 0 \end{aligned}$$

Hence the real and imaginary parts are each zero. Since $\operatorname{Re}(r - is)^j = \operatorname{Re}(r + is)^j$ and $\operatorname{Im}(r - is)^j = -\operatorname{Im}(r + is)^j$ it follows that $p(r - is) = 0$.

In the case that the coefficients are real it may be of interest to note what statement of factorization analogous to (4) above. To this end we need the definition

Definition 3.8.1. The real polynomial $x^2 + bx + c$ is called **irreducible** if it has no real zeros.

In general, any polynomial that cannot be factored over the underlying field is called irreducible. Irreducible polynomials have played a very important role in fields such as abstract algebra and number theory. Indeed, they were central in early attempts to prove Fermat's last theorem and also so but more indirectly to the actual proof. Our attention here is restricted to the fields $\mathbb{C}$ and R .

Theorem 3.8.2. Let $p(\lambda) = \lambda^n + p_1 \lambda^{n-1} + \cdots + p_n$ with coefficients in $p_0, \dots, p_n \in R$. Then $p(\lambda)$ can be factored as a product of linear factors pertaining to real zeros of $p(\lambda) = 0$ and irreducible quadratic factors.

Proof. The proof is an easy application of the FTA and the observation above. If $\lambda_k = r + is$ is a zero of $p(\lambda)$, then so also is $\bar{\lambda}_k = r - is$. Therefore the product $(\lambda - \lambda_k)(\lambda - \bar{\lambda}_k) = \lambda^2 - 2\lambda r + r^2 + s^2$ is an irreducible quadratic. Combine such terms with the linear factors generated from the real zeros, and use (4). $\square$

Remark 3.8.1. It is worth noting that there are **no** higher order irreducible factors with real coefficients. Even though there are certainly higher order polynomials with complex roots, they can always be factored as products of either real linear or real quadratic factors. Proving this without the FTA may prove challenging.

3.9 Appendix B

3.9.1 Infinite Series

Definition 3.9.1. An infinite series, denoted by

$$a_0 + a_1 + a_2 + \cdots$$

is a sequence $\{u_n\}$, where u_n is defined by

$$u_n = a_0 + a_1 + \cdots + a_n$$

If the sequence $\{u_n\}$ converges to some limit A , we say that the infinite series converges to A and use the notation

$$a_0 + a_1 + a_2 + \cdots = A$$

We also say that the sum of the infinite series is A . If $\{u_n\}$ diverges, the infinite series $a_0 + a_1 + a_2 + \cdots$ is said to be divergent.

The sequence $\{u_n\}$ is called the sequence of partial sums, and the sequence $\{a_n\}$ is called the sequence of terms of the infinite series $a_0 + a_1 + a_2 + \cdots$.

Let us now return to the formula

$$1 + r + r^2 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r}$$

where $r \neq 1$. Since the sequence $\{r^{n+1}\}$ converges (and, in fact, to 0) if and only if $-1 < r < 1$ (or $|r| < 1$), r being different from 1, the infinite series

$$1 + r + r^2 + \cdots$$

converges to $(1 - r)^{-1}$ if and only if $|r| < 1$. This infinite series is called the geometric series with ratio r .

Definition 3.9.2. Geometric Series with Ratio r

$$1 + r + r^2 + \cdots = \frac{1}{1 - r}, \quad \text{if } |r| < 1$$

and diverges if $|r| \geq 1$.

Multiplying both sides by r yields the following.

Definition 3.9.3. If $|r| < 1$, then

$$r + r^2 + r^3 + \cdots = \frac{r}{1-r}$$

Example 3.9.1. Investigate the convergence of each of the following infinite series. If it converges, determine its limit (sum).

a. $1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \cdots$ b. $1 + \frac{2}{3} + \frac{4}{9} + \frac{8}{27} + \cdots$ c. $\frac{3}{4} + \frac{9}{16} + \frac{27}{64} + \cdots$

Solution A careful study reveals that each infinite series is a geometric series. In fact, the ratios are, respectively, (a) $-\frac{1}{2}$, (b) $\frac{2}{3}$, and (c) $\frac{3}{4}$. Since they are all of absolute value less than 1, they all converge. In fact, we have:

a. $1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \cdots = \frac{1}{1-(-\frac{1}{2})} = \frac{2}{3}$
 b. $1 + \frac{2}{3} + \frac{4}{9} + \frac{8}{27} + \cdots = \frac{1}{1-\frac{2}{3}} = \frac{3}{1} = 3$
 c. $\frac{3}{4} + \frac{9}{16} + \frac{27}{64} + \cdots = \frac{\frac{3}{4}}{1-\frac{3}{4}} = \frac{3}{4} \cdot \frac{4}{1} = 3$

Since an infinite series is defined as a sequence of partial sums, the properties of sequences carry over to infinite series. The following two properties are especially important.

1. Uniqueness of Limit If $a_0 + a_1 + a_2 + \cdots$ converges, then it converges to a unique limit.

This property explains the notation

$$a_0 + a_1 + a_2 + \cdots = A$$

where A is the limit of the infinite series. Because of this notation, we often say that the infinite series sums to A , or the sum of the infinitely many terms $a_0, a_1, a_2, \dots$ is A . We remark, however, that the order of the terms cannot be changed arbitrarily in general.

Definition 3.9.4. 2. Sum of Infinite Series If

$a_0 + a_1 + a_2 + \cdots = A$ and $b_0 + b_1 + b_2 + \cdots = B$, then

$$\begin{aligned} (a_0 + b_0) + (a_1 + b_1) + (a_2 + b_2) + \cdots &= (a_0 + a_1 + a_2 + \cdots) \\ &\quad + (b_0 + b_1 + b_2 + \cdots) \\ &= A + B \end{aligned}$$

This property follows from the observation that

$$\begin{aligned} (a_0 + b_0) + (a_1 + b_1) + \cdots + (a_n + b_n) &= (a_0 + a_1 + \cdots + a_n) \\ &\quad + (b_0 + b_1 + \cdots + b_n) \end{aligned}$$

converges to $A + B$. Another property is

Definition 3.9.5. 3. Constant Multiple of Infinite Series If $a_0 + a_1 + a_2 + \cdots = A$ and c is a constant, then

$$ca_0 + ca_1 + ca_2 + \cdots = cA$$

Example 3.9.2. Determine the sums of the following convergent infinite series.

a. $\frac{5}{3 \cdot 2} + \frac{13}{9 \cdot 4} + \frac{35}{27 \cdot 8} + \cdots$ b. $\frac{1}{3 \cdot 2} + \frac{1}{9 \cdot 2} + \frac{1}{27 \cdot 2} + \cdots$

Solution a. By the Sum Property, the sum of the first infinite series is

$$\begin{aligned} & \left(\frac{1}{3} + \frac{1}{2} \right) + \left(\frac{1}{9} + \frac{1}{4} \right) + \left(\frac{1}{27} + \frac{1}{8} \right) + \cdots \\ &= \left(\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots \right) + \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \right) \\ &= \frac{\frac{1}{3}}{1 - \frac{1}{3}} + \frac{\frac{1}{2}}{1 - \frac{1}{2}} = \frac{1}{2} + 1 = \frac{3}{2} \end{aligned}$$

b. By the Constant-Multiple Property, the sum of the second infinite series is

$$\frac{1}{2} \left(\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots \right) = \frac{1}{2} \cdot \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Another type of infinite series can be illustrated by using Taylor polynomial extrapolation as follows.

Example 3.9.3. In extrapolating the value of $\ln 2$, the n th-degree Taylor polynomial $P_n(x)$ of $\ln(1+x)$ at $x=0$ was evaluated at $x=1$ in Example 4 of Section 12.3. Interpret $\ln 2$ as the sum of an infinite series whose n th partial sum is $P_n(1)$.

Solution We have seen in Example 3 of Section 12.3 that

$$P_n(x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n-1} \frac{1}{n}x^n$$

so that $P_n(1) = 1 - \frac{1}{2} + \frac{1}{3} - \cdots + (-1)^{n-1} \frac{1}{n}$, and it is the n th partial sum of the infinite series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

Since $\{P_n(1)\}$ converges to $\ln 2$ (see Example 4 in Section 12.3), we have

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

It should be noted that the terms of the infinite series in the preceding example alternate in sign. In general, an infinite series of this type always converges.

Alternating Series Test Let $a_0 \geq a_1 \geq \cdots \geq 0$. If a_n approaches 0, then the infinite series

$$a_0 - a_1 + a_2 - a_3 + \cdots$$

converges to some limit A , where $0 < A < a_0$.

The condition that the term a_n tends to zero is essential in the above Alternating Series Test. In fact, if the terms do not tend to zero, the corresponding infinite series, alternating or not, must diverge.

Divergence Test If a_n does not approach zero as $n \rightarrow +\infty$, then the infinite series

$$a_0 + a_1 + a_2 + \cdots$$

diverges.

Example 3.9.4. Determine the convergence or divergence of the following infinite series.

a. $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$ b. $-\frac{2}{3} + \frac{4}{5} - \frac{6}{7} + \cdots$

Solution a. This series is an alternating series with

$$1 > \frac{1}{3} > \frac{1}{5} > \cdots > 0$$

and with the terms approaching 0. In fact, the general (n th) term is

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

Hence, by the Alternating Series Test, the infinite series $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$ converges.

b. Let $a_1 = -\frac{2}{3}, a_2 = \frac{4}{5}, a_3 = -\frac{6}{7}, \dots$. It is clear that a_n does not approach 0, and it follows from the Divergence Test that the infinite series is divergent. [The n th term is $(-1)^n 2n/(2n+1)$.] Note, however, that the terms do alternate in sign.

Another useful tool for testing convergence or divergence of an infinite series is the Ratio Test, to be discussed below. Recall that the geometric series

$$1 + r + r^2 + \cdots$$

where the n th term is $a_n = r^n$, converges if $|r| < 1$ and diverges otherwise. Note also that the ratio of two consecutive terms is

$$\frac{a_{n+1}}{a_n} = \frac{r^{n+1}}{r^n} = r$$

Hence, the geometric series converges if and only if this ratio is of absolute value less than 1. In general, it is possible to draw a similar conclusion if the sequence of the absolute values of the ratios of consecutive terms converges.

Ratio Test Suppose that the sequence

$$\{|a_{n+1}|/|a_n|\}$$

converges to some limit R . Then, the infinite series $a_0 + a_1 + a_2 + \cdots$ converges if $R < 1$ and diverges if $R > 1$.

Example 3.9.5. In each of the following, determine all values of r for which the infinite series converges.

a. $1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} + \cdots$ b. $1 + 2r + 3r^3 + 4r^3 + \cdots$

Solution a. The n th term is $a_n = r^n/n!$, so that

$$\frac{a_{n+1}}{a_n} = \frac{r^{n+1}}{(n+1)!} \cdot \frac{n!}{r^n} = \frac{r}{n+1}$$

For each value of r , we have

$$\frac{|a_{n+1}|}{|a_n|} = \frac{1}{n+1}|r| \rightarrow 0 < 1$$

Hence, by the Ratio Test, the infinite series converges for all values of r .

b. Let $a_n = (n+1)r^n$. Then,

$$\frac{|a_{n+1}|}{|a_n|} = \frac{(n+2)|r|^{n+1}}{(n+1)|r|^n} = \frac{n+2}{n+1}|r| \rightarrow |r|$$

Hence, the Ratio Test says that the infinite series converges if $|r| < 1$ and diverges if $|r| > 1$. For $|r| = 1$, we have $|a_n| = n+1$, which does not converge to 0, so that the corresponding infinite series must be divergent as a result of applying the Divergence Test.

Sometimes it is possible to compare the partial sums of an infinite series with certain integrals, as illustrated in the following example.

Example 3.9.6. Show that $1 + \frac{1}{2} + \cdots + 1/n$ is larger than the definite integral $\int_1^{n+1} dx/x$, and conclude that the infinite series $1 + \frac{1}{2} + \frac{1}{3} + \cdots$ diverges.

Solution Consider the function $f(x) = 1/x$. Then, the definite integral

$$\int_1^{n+1} \frac{dx}{x} = \ln(n+1) - \ln 1 = \ln(n+1)$$

is the area under the graph of the function $y = f(x)$ between $x = 1$ and $x = n+1$. On the other hand, the sum

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} = [f(1) + f(2) + \cdots + f(n)]\Delta x$$

where $\Delta x = 1$, is the sum of the areas of n rectangles with base $\Delta x = 1$ and heights $f(1), \dots, f(n)$, consecutively, as shown in Figure 12.5. Since the union of these rectangles covers the region bounded by the curve $y = f(x)$, the x -axis, and the vertical lines $x = 1$ and $x = n+1$, we have:

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} > \int_1^{n+1} \frac{dx}{x} = \ln(n+1)$$

Now recall that $\ln(n+1)$ approaches ∞ as n approaches ∞ . Hence, the sequence of partial sums of the infinite series

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

must be divergent.

The above infinite series is called the harmonic series. It diverges “to infinity” in the sense that its sequence of partial sums becomes arbitrarily large for all large values of n . If an infinite series diverges to infinity, we also say that it sums to infinity and use the notation “ $= \infty$ ” accordingly. We have

The **Harmonic Series** is defined by

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots = \infty$$

Chapter 4

Unitary Matrices

4.1 Basics

This chapter considers a very important class of matrices that are quite useful in proving a number of structure theorems about all matrices. Called *unitary* matrices, they comprise a class of matrices that have the remarkable properties that as transformations they preserve length, and preserve the angle between vectors. This is of course true for the identity transformation. Therefore it is helpful to regard unitary matrices as “generalized identities,” though we will see that they form quite a large class. An important example of these matrices, the rotations, have already been considered. In this chapter, the underlying field is usually $\mathbb{C}$, the underlying vector space is $\mathbb{C}_n$, and almost without exception the underlying norm is $\|\cdot\|_2$. We begin by recalling a few important facts.

Recall that a set of vectors $x_1, \dots, x_k \in \mathbb{C}_n$ is called **orthogonal** if $x_j^* x_m = \langle x_m, x_j \rangle = 0$ for $1 \leq j \neq m \leq k$. The set is **orthonormal** if

$$x_j^* x_m = \delta_{mj} = \begin{cases} 1 & j = m \\ 0 & j \neq m. \end{cases}$$

An orthogonal set of vectors can be made orthonormal by scaling:

$$x_j \longrightarrow \frac{1}{(x_j^* x_j)^{1/2}} x_j.$$

Theorem 4.1.1. *Every set of orthonormal vectors is linearly independent.*

Proof. The proof is routine, using a common technique. Suppose $S = \{x_j\}_{j=1}^k$ is orthonormal and linearly dependent. Then, without loss of gen-

erality (by relabeling if needed), we can assume

$$u_k = \sum_{j=1}^{k-1} c_j u_j.$$

Compute $u_k^* u_k$ as

$$\begin{aligned} 1 = u_k^* u_k &= u_k^* \left(\sum_{j=1}^{k-1} c_j u_j \right) \\ &= \sum_{j=1}^{k-1} c_j u_k^* u_j = 0. \end{aligned}$$

This contradiction proves the result. $\square$

Corollary 4.1.1. *If $S = \{u_1 \dots u_k\} \subset \mathbb{C}_n$ is orthonormal then $k \leq n$.*

Corollary 4.1.2. *Every k -dimensional subspace of $\mathbb{C}_n$ has an orthonormal basis.*

Proof. Apply the Gram–Schmidt process to any basis to orthonormalize it. $\square$

Definition 4.1.1. A matrix $U \in M_n$ is said to be **unitary** if $U^* U = I$. [If $U \in M_n(\mathbb{R})$ and $U^T U = I$, then U is called real **orthogonal**.]

Note: A linear transformation $T: \mathbb{C}_n \rightarrow \mathbb{C}_n$ is called an **isometry** if $\|Tx\| = \|x\|$ for all $x \in \mathbb{C}_n$.

Proposition 4.1.1. *Suppose that $U \in M_n$ is unitary. (i) Then the columns of U form an orthonormal basis of $\mathbb{C}_n$, or $\mathbb{R}_n$, if U is real. (ii) The spectrum $\sigma(u) \subset \{z \mid |z| = 1\}$. (iii) $|\det U| = 1$.*

Proof. The proof of (i) is a consequence of the definition. To prove (ii), first denote the columns of U by u_i , $i = 1, \dots, n$. If λ is an eigenvalue of U with pertaining eigenvector x , then $\|Ux\| = \|\sum x_i u_i\| = (\sum |x_i|^2)^{1/2} = \|x\| = |\lambda| \|x\|$. Hence $|\lambda| = 1$. Finally, (iii) follows directly because $\det U = \prod \lambda_i$. Thus $|\det U| = \prod |\lambda_i| = 1$. $\square$

This important result is just one of many equivalent results about unitary matrices. In the result below, a number of equivalences are established.

Theorem 4.1.2. *Let $U \in M_n$. The following are equivalent.*

- (a) U is unitary.
- (b) U is nonsingular and $U^* = U^{-1}$.
- (c) $UU^* = I$.
- (d) U^* is unitary.
- (e) The columns of U form an orthonormal set.
- (f) The rows of U form an orthonormal set.
- (g) U is an isometry.
- (h) U carries every set of orthonormal vectors to a set of orthonormal vectors.

Proof. (a) $\Rightarrow$ (b) follows from the definition of unitary **and** the fact that the inverse is unique.

(b) $\Rightarrow$ (c) follows from the fact that a left inverse is also a right inverse.

(a) $\Rightarrow$ (d) $UU^* = (U^*)^*U^* = I$.

(d) $\Rightarrow$ (e) (e) $\equiv$ (b) $\Rightarrow u_j^* u_k \delta_{jk}$, where $u_1 \dots u_n$ are the columns of U . Similarly (b) $\Rightarrow$ (e).

(d) $\equiv$ (f) same reasoning.

(e) $\Rightarrow$ (g) We know the columns of U are orthonormal. Denoting the columns by $u_1 \dots u_n$, we have

$$Ux = \sum_1^n x_i u_i$$

where $x = (x_1, \dots, x_n)^T$. It is an easy matter to see that

$$\|Ux\|^2 = \sum_1^n |x_i|^2 = \|x\|^2.$$

(g) $\Rightarrow$ (e). Consider $x = e_j$. Then $Ux = u_j$. Hence $1 = \|e_j\| = \|Ue_j\| = \|u_j\|$. The columns of U have norm one. Now let $x = \alpha e_i + \beta e_j$ be chosen

such that $\|x\| = \|\alpha e_i + \beta e_j\| = \sqrt{|\alpha|^2 + |\beta|^2} = 1$. Then

$$\begin{aligned}
 1 &= \|Ux\|^2 \\
 &= \|U(\alpha e_i + \beta e_j)\|^2 \\
 &= \langle U(\alpha e_i + \beta e_j), U(\alpha e_i + \beta e_j) \rangle \\
 &= |\alpha|^2 \langle Ue_i, Ue_i \rangle + |\beta|^2 \langle Ue_j, Ue_j \rangle + \alpha \bar{\beta} \langle Ue_i, Ue_j \rangle + \bar{\alpha} \beta \langle Ue_j, Ue_i \rangle \\
 &= |\alpha|^2 \langle u_i, u_i \rangle + |\beta|^2 \langle u_j, u_j \rangle + \alpha \bar{\beta} \langle u_i, u_j \rangle + \bar{\alpha} \beta \langle u_j, u_i \rangle \\
 &= |\alpha|^2 + |\beta|^2 + 2\Re(\alpha \bar{\beta} \langle u_i, u_j \rangle) \\
 &= 1 + 2\Re(\alpha \bar{\beta} \langle u_i, u_j \rangle)
 \end{aligned}$$

Thus $\Re(\alpha \bar{\beta} \langle u_i, u_j \rangle) = 0$. Now suppose $\langle u_i, u_j \rangle = s + it$. Selecting $\alpha = \beta = \frac{1}{\sqrt{2}}$ we obtain that $\Re \langle u_i, u_j \rangle = 0$, and selecting $\alpha = i\beta = \frac{1}{\sqrt{2}}$ we obtain that $\Im \langle u_i, u_j \rangle = 0$. Thus $\langle u_i, u_j \rangle = 0$. Since the coordinates i and j are arbitrary, it follows that the columns of U are orthogonal.

(g) $\Rightarrow$ (h) Suppose $\{v_1, \dots, v_n\}$ is orthogonal. For any two of them $\|U(v_j + v_k)\|^2 = \|v_j + v_k\|^2$. Hence $\langle Uv_j, Uv_k \rangle = 0$.

(h) $\Rightarrow$ (e) The orthonormal set of the standard unit vectors e_j , $j = 1, \dots, n$ is carried to the columns of U . That is $Ue_j = u_j$, the j^{th} column of U . Therefore the columns of U are orthonormal. $\square$

Corollary 4.1.3. *If $U \in M_n(\mathbb{C})$ is unitary, then the transformation defined by U preserves angles.*

Proof. We have for any vectors $x, y \in \mathbb{C}_n$ that the angle θ is completely determined from the inner product via $\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$. Since U is unitary (and thus an isometry) it follows that

$$\langle Ux, Uy \rangle = \langle U^* Ux, y \rangle = \langle x, y \rangle$$

This proves the result. $\square$

Example 4.1.1. Let $T(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ where θ is any real. Then $T(\theta)$ is real orthogonal.

Proposition 4.1.1. *If $U \in M_2(\mathbb{R})$ is real orthogonal, then U has the form $T(\theta)$ for some θ or the form*

$$U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} T(\theta) = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$$

Finally, we can easily establish the diagonalizability of unitary matrices.

Theorem 4.1.3. *If $U \in M_n$ is unitary, then it is diagonalizable.*

Proof. To prove this we need to revisit the proof of Theorem 3.5.2. As before, select the first vector to be a normalized eigenvector u_1 pertaining to λ_1 . Now choose the remaining vectors to be orthonormal to u_1 . This makes the matrix P_1 with all these vectors as columns a unitary matrix. Therefore $B_1 = P^{-1}UP$ is also unitary. However it has the form

$$B_1 = \begin{bmatrix} \lambda & \alpha_1 & \dots & \alpha_{n-1} \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix} \quad \text{where } A_2 \text{ is } (n-1) \times (n-1)$$

Since it is unitary, it must have orthogonal columns by Theorem 4.1.2. It follows then that $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ and

$$B_1 = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix}$$

At this point one may apply an inductive hypothesis to conclude that A_2 is similar to a diagonal matrix. Thus by the manner in which the full similarity was constructed, we see that A must also be similar to a diagonal matrix. $\square$

Corollary 4.1.1. *Let $U \in M_n$ be unitary. Then*

- (i) *Then U has a set of n orthogonal eigenvectors.*
- (ii) *Let $\{\lambda_1, \dots, \lambda_n\}$ and $\{v_1, \dots, v_n\}$ denote respectively the eigenvalues and their pertaining orthonormal eigenvectors of U . Then U has the representation as the sum of rank one matrices given by*

$$U = \sum_{j=1}^n \lambda_j v_j v_j^T$$

*This representation is often called the **spectral representation** or **spectral decomposition** of U .*

4.1.1 Groups of matrices

Invertible and unitary matrices have a fundamental structure that makes possible a great many general statements about their nature and the way they act upon vectors other vectors matrices. A group is a set with a mathematical operation, product, that obeys some minimal set of properties so as to resemble the nonzero numbers under multiplication.

Definition 4.1.2. A **group** G is a set with a binary operation $G \times G \rightarrow G$ which assigns to every pair a, b of elements of G a unique element ab in G . The operation, called the **product**, satisfies four properties:

1. Closure. If $a, b \in G$, then $ab \in G$.
2. Associativity. If $a, b, c \in G$, then $a(bc) = (ab)c$.
3. Identity. There exists an element $e \in G$ such that $ae = ea = a$ for every $a \in G$. e is called the *identity* of G .
4. Inverse. For each $a \in G$, there exists an element $\hat{a} \in G$ such that $a\hat{a} = \hat{a}a = e$. $\hat{a}$ is called the *inverse* of a and is often denoted by a^{-1} .

A subset of G that is itself a group under the same product is called a **subgroup** of G .

It may be interesting to note that removal of any of the properties 2-4 leads to other categories of sets that have interest, and in fact applications, in their own right. Moreover, many groups have additional properties such as commutativity, i.e. $ab = ba$ for all $a, b \in G$. Below are a few examples of matrix groups. Note matrix addition is not involved in these definitions.

Example 4.1.2. As usual M_n is the vector space of $n \times n$ matrices. The product in these examples is the usual matrix product.

- The group $GL(n, F)$ is the group of invertible $n \times n$ matrices. This is the so-called **general linear** group. The subset of M_n of invertible lower (resp. upper) triangular matrices is a subgroup of $GL(n, F)$.
- The **unitary** group U_n of unitary matrices in $M_n(\mathbb{C})$.
- The **orthogonal** group O_n orthogonal matrices in $M_n(R)$. The subgroup of O_n denoted by SO_n consists of orthogonal matrices with determinant 1.

Because element inverses are required, it is obvious that the only subsets of invertible matrices in M_n will be groups. Clearly, $GL(n, F)$ is a group because the properties follow from those matrix of multiplication. We have already established that invertible lower triangular matrices have lower triangular inverses. Therefore, they form a subgroup of $GL(n, F)$. We consider the unitary and orthogonal groups below.

Proposition 4.1.2. *For any integer $n = 1, 2, \dots$ the set of unitary matrices U_n (resp. real orthogonal) forms a group. Similarly O_n is a group, with subgroup SO_n .*

Proof. The result follows if we can show that unitary matrices are closed under multiplication. Let U and V be unitary. Then

$$\begin{aligned}(UV)^*(UV) &= V^*U^*UV \\ &= V^*V = I\end{aligned}$$

For orthogonal matrices the proof is essentially identical. That SO_n is a group follows from the determinant equality $\det(AB) = \det A \det B$. Therefore it is a subgroup of O_n . $\square$

4.1.2 Permutation matrices

Another example of matrix groups comes from the idea of permutations of integers.

Definition 4.1.3. The matrix $P \in M_n(\mathbb{C})$ is called a **permutation matrix** if each row and each column has exactly one 1, the rest of the entries being zero.

Example 4.1.3. Let

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

P and Q are permutation matrices.

Another way to view a permutation matrix is with the game of chess. On an $n \times n$ chess board place n rooks in positions where *none* of them attack one another. Viewing the board as an $n \times n$ matrix with ones where the rooks are and zeros elsewhere, this matrix will be a permutation matrix.

Of course there are $n!$ such placements, exactly the number of permutations of the integers $\{1, 2, \dots, n\}$.

Permutation matrices are closely linked with permutations as discussed in Chapter 2.5. Let σ be a permutation of the integers $\{1, 2, \dots, n\}$. Define the matrix A by

$$a_{ij} = \begin{cases} 1 & \text{if } j = \sigma(i) \\ 0 & \text{if otherwise} \end{cases}$$

Then A is a permutation matrix. We could also use the Dirac notation to express the same matrix, that is to say $a_{ij} = \delta_{i\sigma(j)}$. The product of permutation matrices is again a permutation matrix. This is apparent by straight multiplication. Let P and Q be two $n \times n$ permutation matrices with pertaining permutations σ_P and σ_Q of the integers $\{1, 2, \dots, n\}$. Then the i^{th} row of P is $e_{\sigma_P(i)}$ and the i^{th} row of Q is $e_{\sigma_Q(i)}$. (Recall the e_i are the usual standard vectors.) Now the i^{th} row of the product PQ can be computed by

$$\sum_{j=1}^n p_{ij} e_{\sigma_Q(j)} = \sum_{j=1}^n \delta_{i\sigma(j)} e_{\sigma_Q(j)} = e_{\sigma_Q(\sigma_P(i))}$$

Thus the multiplication i^{th} row of PQ is a standard vector. Since the $\sigma_P(i)$ ranges over the integers $\{1, 2, \dots, n\}$, it is true also that $\sigma_Q(\sigma_P(i))$ does likewise. Therefore the “product” $\sigma_Q(\sigma_P(i))$ is also a permutation. We conclude that the standard vectors constitute the rows of PQ . Thus permutation matrices are orthogonal under multiplication. Moreover the inverse of every permutation is permutation matrix, with the inverse describe through the inverse of the pertaining permutation of $\{1, 2, \dots, n\}$. Therefore, we have the following result.

Proposition 4.1.3. *Permutation matrices are orthogonal. Permutation matrices form a subgroup of O_n .*

4.1.3 Unitary equivalence

Definition 4.1.4. A matrix $B \in M_n$ is said to be *unitarily equivalent* to A if there is a unitary matrix $U \in M_n$ such that

$$B = U^*AU.$$

(In the real case we say B is orthogonally equivalent to A .)

Theorem 4.1.4. *If B and A are unitarily equivalent. Then*

$$\sum_{i,j=1}^n |b_{ij}|^2 = \sum_{i,j=1}^n |a_{ij}|^2.$$

Proof. We have $\sum_{i,j=1}^n |a_{ij}|^2 = \text{tr } A^*A$, and

$$\begin{aligned} \sum_{i,j=1}^n |b_{ij}|^2 &= \text{tr } B^*B = \text{tr } (U^*AU)^*U^*AU = \text{tr } U^*A^*AU \\ &= \text{tr } A^*A, \end{aligned}$$

since the trace is invariant under similarity transformations. $\square$

Alternatively, we have $BU^* = U^*A$. Since U (and U^*) are isometries we have each column of U^*A has the same norm as the norm of the same column of A . The same holds for the rows of BU^* , whence the result.

Example 4.1.4. $B = \begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ are similar but not unitarily equivalent. A and B are similar because (1) they have the same spectrum, $\sigma(A) = \sigma(B) = \{1, 2\}$ and (2) they have two linearly independent eigenvectors. They are not unitarily equivalent because the conditions of the above theorem are not met.

Remark 4.1.1. Unitary equivalence is a finer classification than similarity. Indeed, consider the two sets $S(A) = \{B \mid B \text{ is similar to } A\}$ and $U(A) = \{B \mid B \text{ is unitarily equivalent to } A\}$, then

$$U(A) \subset S(A).$$

(Can you show that $U(A) \subsetneq S(A)$ for some large class of $A \in M_n$?)

4.1.4 Householder transformations

An important class of unitary transformations are elementary reflections. These can be realized as transformations that reflect one vector to its negative and leave invariant the orthocomplement of vectors.

Definition 4.1.5. Simple Householder transformation.) Suppose $w \in \mathbb{C}^n$, $\|w\| = 1$. Define the **Householder** transformation H_w by

$$H_w = I - 2ww^*$$

Computing

$$\begin{aligned}
 H_w H_w^* &= (I - 2ww^*)(I - 2ww^*)^* \\
 &= I - 2ww^* - 2ww^* + 4ww^*ww^* \\
 &= I - 4ww^* + 4\langle w, w \rangle ww^* = I
 \end{aligned}$$

it follows that H_w is unitary.

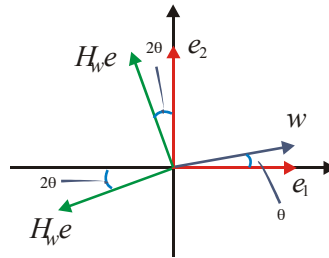
Example 4.1.5. Consider the vector $w = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$ and the Householder transformation

$$\begin{aligned}
 H_w &= I - 2ww^T = \begin{bmatrix} 1 - 2\cos^2 \theta & -2\cos \theta \sin \theta \\ -2\cos \theta \sin \theta & 1 - 2\sin^2 \theta \end{bmatrix} \\
 &= \begin{bmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix}
 \end{aligned}$$

The transformation properties for the standard vectors are

$$\begin{aligned}
 H_w e_1 &= \begin{bmatrix} -\cos 2\theta \\ -\sin 2\theta \end{bmatrix} \quad \text{and} \\
 H_w e_2 &= \begin{bmatrix} \cos 2\theta \\ -\sin 2\theta \end{bmatrix}
 \end{aligned}$$

This is shown below. It is evident that this unitary transformation is not a rotation. Though, it can be imagined as a “rotation with a one dimensional reflection.”



Householder transformation H_w

Example 4.1.6. Let θ be real. For any $n \geq 2$ and $1 \leq i, j \leq n$ with $i \neq j$

define

$$U_n(\theta, i, j) = \begin{bmatrix} 1 & 0 & \cdots & & \vdots & & 0 \\ & \ddots & & \vdots & & & \\ 0 & 1 & \vdots & & \vdots & & \\ \cdots & \cdots & \cos \theta & \cdots & \cdots & -\sin \theta & \cdots \\ & & & 1 & & & \\ & & \vdots & & \ddots & & \vdots \\ & & & & & 1 & \\ \cdots & \cdots & \sin \theta & \cdots & \cdots & \cos \theta & \cdots \\ & & \vdots & & & \vdots & 1 \\ & & \vdots & & & \vdots & 0 \\ & & \vdots & & & \vdots & \ddots \\ 0 & & \vdots & & & \vdots & 0 \\ & & & i & & & j \\ & & & & & & 1 \end{bmatrix}$$

Then $U_n(\theta; i, j)$ is a rotation and is unitary.

Proposition 4.1.4 (Limit theorems). (i) Show that the unitary matrices are closed with respect to any norm. That is, if the sequence $\{U_n\} \subset M_n(\mathbb{C})$ are all unitary and the $\lim_{n \rightarrow \infty} U_n = U$ in the $\|\cdot\|_2$ norm, then U is also unitary.

(ii) The unitary matrices are closed under pointwise convergence. That is, if the sequence $\{U_n\} \subset M_n(\mathbb{C})$ are all unitary and $\lim_{n \rightarrow \infty} U_n = U$ for each (ij) entry, then U is also unitary.

Householder transformations can also be used to triangularize a matrix. The procedure successively removes the lower triangular portion of a matrix column by column in a way similar to Gaussian elimination. The elementary row operations are replaced by elementary reflectors. The result is a triangular matrix $T = H_{v_n} \cdots H_{v_2} H_{v_1} A$. Since these reflectors are unitary, the factorization yields an effective method for solving linear systems. The actual process is rather straightforward. Construct the vector v such that

$$(I - 2vv^T) A = (I - 2vv^T) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ 0 & \hat{a}_{22} & \cdots & \hat{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{a}_{n2} & \cdots & \hat{a}_{nn} \end{bmatrix}$$

This is accomplished as follows. This means we wish to find a vector v such that

$$(I - 2vv^T) [a_{11}, a_{21}, \dots, a_{n1}]^T = [\hat{a}_{11}, 0, \dots, 0]^T$$

For notational convenience and to emphasize the construction is vector based, relabel the column vector $[a_{11}, a_{21}, \dots, a_{n1}]^T$ as $[x_1, x_2, \dots, x_n]^T$

$$v_j = \frac{x_j}{2\langle v, x \rangle}$$

for $j = 2, 3, \dots, n$. Define

$$v_1 = \frac{x_1 + \alpha}{2\langle v, x \rangle}$$

Then

$$\begin{aligned} \langle v, x \rangle &= \frac{\langle x, x \rangle}{2\langle v, x \rangle} + \frac{\alpha x_1}{2\langle v, x \rangle} = \frac{\|x\|^2}{2\langle v, x \rangle} + \frac{\alpha x_1}{2\langle v, x \rangle} \\ 4\langle v, x \rangle^2 &= 2\|x\|^2 + 2\alpha x_1 \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm. Also, for H_v to be unitary we need

$$\begin{aligned} 1 &= \langle v, v \rangle = \frac{1}{4\langle v, x \rangle^2} (\|x\|^2 + 2\alpha x_1 + \alpha^2) \\ 4\langle v, x \rangle^2 &= \|x\|^2 + 2\alpha x_1 + \alpha^2 \end{aligned}$$

Equating the two expressions for $4\langle v, x \rangle^2$ gives

$$\begin{aligned} 2\|x\|^2 + 2\alpha x_1 &= \|x\|^2 + 2\alpha x_1 + \alpha^2 \\ \alpha^2 &= \|x\|^2 \\ \alpha &= \pm \|x\| \end{aligned}$$

We now have that

$$\begin{aligned} 4\langle v, x \rangle^2 &= 2\|x\|^2 \pm 2\|x\| x_1 \\ \langle v, x \rangle &= \left(\frac{1}{2} (\|x\|^2 \pm \|x\| x_1) \right)^{\frac{1}{2}} \end{aligned}$$

This makes $v_1 = \frac{x_1 \pm \|x\|}{\left(\frac{1}{2} (\|x\|^2 \pm \|x\| x_1) \right)^{\frac{1}{2}}}$. With the construction of H_v to “eliminate” the first column of A , we relabel the vector v as v_1 (with the small

possibility of notational confusion) and move to describe H_{v_2} using a transformation of the same kind with a vector of the type $v = (0, x_2, x_3, \dots, x_n)$. Such a selection will not affect the structure of the first column. Continue this until the matrix is triangularized. The upshot is that every matrix can be factored as $A = UT$, where U is unitary and T is upper triangular.

The main result for this section is the factorization theorem. As it turns out, every Unitary matrix can be written as a product of elementary reflectors. The proof requires a slightly more general notion of reflector.

Definition 4.1.6. The general form of the Householder matrix, also called an **elementary reflector**, has the form

$$H_v = I - \tau v v^*$$

where the vector $v \in C_n$.

In order for H_v to be unitary, it must be true that

$$\begin{aligned} H_v H_v^* &= (I - \tau v v^*) (I - \tau v v^*)^* \\ &= (I - \tau v v^*) (I - \bar{\tau} v v^*) \\ &= I - \tau v v^* - \bar{\tau} v v^* + |\tau|^2 (v^* v) v v^* \\ &= I - 2\operatorname{Re}(\tau) v v^* + |\tau|^2 |v|^2 v v^* \\ &= I \end{aligned}$$

Therefore, for $v \neq 0$ we must have

$$\begin{aligned} -2\operatorname{Re}(\tau) v v^* + |\tau|^2 |v|^2 v v^* &= \left(-2\operatorname{Re}(\tau) + |\tau|^2 |v|^2 \right) v v^* \\ &= 0 \end{aligned}$$

or

$$-2\operatorname{Re}(\tau) + |\tau|^2 |v|^2 = 0$$

Now suppose that $Q \in M_n(R)$ is orthogonal and that the spectrum $\sigma(Q) \subset \{-1, 1\}$. Suppose Q has a complete set of normalized orthogonal eigenvectors¹, it can be expressed as $Q = \sum_{i=1}^n \lambda_i v_i v_i^T$ where the set $v_1, \dots, v_n$ are the eigenvectors and $\lambda_i \in \sigma(Q)$. Now assume the eigenvectors have been arranged so that this simplifies to

$$Q = - \sum_{i=1}^k v_i v_i^T + \sum_{i=k+1}^n v_i v_i^T$$

¹This is in fact a theorem that will be established in Chapter 4.2. It follows as a consequence of Schur's theorem

Here we have just arrange the eigenvectors with eigenvalue -1 to come first. Define

$$H_j = I - 2v_j v_j^T, \quad j = 1, \dots, k$$

It follows that

$$Q = \prod_{j=1}^k H_j = \prod_{j=1}^k (I - 2v_j v_j^T)$$

for it is easy to check that

$$Uv_m = \prod_{j=1}^k H_j v_m = \begin{cases} -v_m & \text{if } m \leq k \\ v_m & \text{if } m > k \end{cases}$$

Since we have agreement with Q on a basis, the equality follows. In words we may say that an orthogonal matrix U with spectrum $\sigma(Q) \subset \{-1, 1\}$ can be written as a product of reflectors. With this simple case out of the way, we consider more general case we write $Q = \sum_{i=1}^n \lambda_i v_i v_i^T$. where the set $v_1, \dots, v_n$ are the eigenvectors and $\lambda_i \in \sigma(Q)$. We wish to represent Q similar to the above formula as a product of elementary reflectors.

$$Q = \prod_{j=1}^k H_j = \prod_{j=1}^k (I - \tau_j w_j w_j^*)$$

where $w_i = \alpha_i v_i$ for some scalars α_i . On the one hand it must be true that $-2\operatorname{Re}(\tau_i) + |\tau_i|^2 |w_i|^2 = 0$ for each $i = 1, \dots, n$, and on the other hand it must follow that

$$(I - \tau_i w_i w_i^*) v_m = \begin{cases} \lambda_i v_i & \text{if } m = i \\ v_m & \text{if } m \neq i \end{cases}$$

The second of these relations is automatically satisfied by the orthogonality of the eigenvectors. The second relation can needs to be solved. This simplifies to $(I - \tau_i w_i w_i^*) v_i = (1 - \tau_i |\alpha_i|^2) v_i = \lambda_i v_i$. Therefore, it is necessary to solve the system

$$\begin{aligned} -2\operatorname{Re}(\tau_i) + |\tau_i|^2 |v_i|^2 &= 0 \\ 1 - \tau_i |\alpha_i|^2 &= \lambda_i \end{aligned}$$

Having done so there results the factorization.

Theorem 4.1.5. *Let $Q \in M_n$ be real orthogonal or unitary. Then Q can be factored as the product of elementary reflectors $Q = \prod_{j=1}^n (I - \tau_j w_j w_j^*)$, where the w_j are the eigenvectors of Q .*

Note that the product written here is up to n whereas the earlier product was just up to k . The difference here is slight, for by taking the scalar (α) equal zero when necessary, it is possible to equate the second form to the first form when the spectrum is contained in the set $\{-1, 1\}$.

4.2 Schur's theorem

It has already been established in Theorem 3.5.2 that every matrix is similar to a triangular matrix. A far stronger result is possible. Called Schur's theorem, this result proves that the similarity is actually unitary similarity.

Theorem 4.2.1 (Schur's Theorem). *Every matrix $A \in M_n(\mathbb{C})$ is unitarily equivalent to a triangular matrix.*

Proof. We proceed by induction on the size of the matrix n . First suppose the A is 2×2 . Then for a given eigenvalue λ and normalized eigenvector v form the matrix P with its first column v and second column any vector orthogonal to v with norm one. Then P is a unitary matrix and $P^*AP = \begin{bmatrix} \lambda & * \\ 0 & * \end{bmatrix}$. This is the desired triangular form. Now assume the Schur factorization is possible for matrices up to size $(n-1) \times (n-1)$. For the given $n \times n$ matrix A select any eigenvalue λ . With its pertaining normalized eigenvector v construct the matrix P with v in the first column and an orthonormal complementary basis in the remaining $n-1$ columns. Then

$$P^*AP = \begin{bmatrix} \lambda & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ 0 & \hat{a}_{22} & \cdots & \hat{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{a}_{n2} & \cdots & \hat{a}_{nn} \end{bmatrix} = \begin{bmatrix} \lambda & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix}$$

The eigenvalues of A_2 together with λ constitute the eigenvalues of A . By the inductive hypothesis there is an $(n-1) \times (n-1)$ unitary matrix $\hat{Q}$ such that $\hat{Q}^*A_2\hat{Q} = T_2$, where T_2 is triangular. Now embed $\hat{Q}$ in an $n \times n$ matrix

Q as shown below

$$Q = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \hat{Q} & \\ 0 & & & \end{bmatrix}$$

It follows that

$$Q^* P^* A P Q = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & T_2 & \\ 0 & & & \end{bmatrix} = T$$

is triangular, as indicated. Therefore the factorization is complete upon defining the similarity transformation $U = PQ$. Of course, the eigenvalues of A_2 together with λ constitute the eigenvalues of A , and therefore by similarity the diagonal of T contains only eigenvalues of A . $\square$

The triangular matrix above is not unique as is easy to see by mixing or permuting the eigenvalues. An important consequence of Schur's theorem pertains to unitary matrices. Suppose that B is unitary and we apply the Schur factorization to write $B = U^* T U$. Then $T = U B U^*$. It follows that the upper triangular matrix is itself unitary, which is to say $T^* T = I$. It is a simple fact to prove this implies that T is in fact a diagonal matrix. Thus, the following result is proved.

Corollary 4.2.1. *Every unitary matrix is diagonalizable. Moreover, every unitary matrix has n orthogonal eigenvectors.*

Proposition 4.2.1. *If $B, A \in M_2$ are similar and $\text{tr}(B^* B) = \text{tr}(A^* A)$, then B and A are unitarily equivalent.*

This result higher dimensions.

Theorem 4.2.2. *Suppose that $A \in M_n(\mathbb{C})$ has distinct eigenvalues $\lambda_1 \dots \lambda_k$ with multiplicities $n_1, n_2, \dots, n_k$ respectively. Then A is similar to a matrix of the form*

$$\begin{bmatrix} T_1 & & & & \\ & & & 0 & \\ & & T_2 & & \\ & 0 & & \ddots & \\ & & & & T_k \end{bmatrix}$$

where $T_i \in M_{n_i}$ is upper triangular with diagonal entries λ_i .

Proof. Let E_1 be eigenspace of λ_1 . Assume $\dim E_1 = m_1$. Let $u_1 \dots u_{m_1}$ be an orthogonal basis of E_1 , and suppose $v_1 \dots v_{n-m_1}$ is an orthonormal basis of $\mathbb{C}^n - E_1$. Then with $P = [u_1 \dots u_{m_1}, v_1 \dots v_{n-m_1}]$ we have $P^{-1}AP$ has the block structure

$$\left[\begin{array}{c|c} T_1 & 0 \\ \hline 0 & A_2 \end{array} \right].$$

Continue this way, as we have done before. However, if $\dim E_i < m_i$ we must proceed a different way. First apply Schur's Theorem to triangularize. Arrange the first m_1 eigenvalues to the first m_1 diagonal entries of T , the similar triangular matrix. Let $E_{r,s}$ be the matrix with a 1 in the r, s position and 0's elsewhere. We assume throughout that $r \neq s$. Then, it is easy to see that $I + \alpha E_{r,s}$ is invertible and that $(I + \alpha E_{r,s})^{-1} = I - \alpha E_{r,s}$. Now if we define

$$(I + \alpha E_{r,s})^{-1} T (I + \alpha E_{r,s})$$

we see that $t_{rs} \rightarrow t_{rs} + \alpha(t_{rr} - t_{ss})$. We know $t_{rr} - t_{ss} \neq 0$ if r and s pertain to values with different eigenvalues. Now this value is changed, but so also are the values above and to the right of t_{rs} . This is illustrated below.

columns $\rightarrow$	s			
		$\uparrow$		
	$*$	$\uparrow$		
row r		$*$	$\rightarrow$	$\rightarrow$

To see how to use these similarity transformation to zero out the upper blocks, consider the special case with just two blocks

$$A = \left[\begin{array}{c|c} T_1 & T_{12} \\ \hline 0 & T_2 \end{array} \right]$$

We will give a procedure to use elementary matrices αE_{ij} to zero-out each of the entries in T_{12} . The matrix T_{12} has $m_1 \cdot m_2$ entries and we will need

to use exactly that many of the matrices αE_{ij} . In the diagram below we illustrate the order of removal (zeroing-out) of upper entries

$$\left[\begin{array}{c|cccc} & m_1 & 2m_1 & \cdots & m_1 \cdot m_2 \\ T_1 & \vdots & \vdots & & \vdots \\ & 2 & \vdots & & \\ \hline & 1 & m_1 + 1 & \cdots & \cdots \\ 0 & T_2 & & & \end{array} \right]$$

When performed in this way we develop $m_1 \cdot m_2$ constants α_{ij} . We proceed in the upper right block (T_{12}) in the left-most column and bottom entry, that is position $(m_1, m_1 + 1)$. For the $(m_1, m_1 + 1)$ position we use the elementary matrix $\alpha E_{m_1, m_1 + 1}$ to zero out this position. This is possible because the diagonal entries of T_1 and T_2 are different. Now use the elementary matrix $\alpha E_{m_1 - 1, m_1 + 1}$ to zero-out the position $(m_1 - 1, m_1 + 1)$. Proceed up this column to the first row each time using a new α . We are finished when all the entries in column $m_1 + 1$ from row m_1 to row 1 are zero. Now focus on the next column to the right, column $m_1 + 2$. Proceed in the same way from the $(m_1, m_1 + 2)$ position to the $(1, m_1 + 1)$ position, using elementary matrices $\alpha E_{k, m_1 + 2}$ zeroing out the entries $(k, m_1 + 2)$ positions, $k = m_1, \dots, 1$. Ultimately, we can zero out every entry in the upper right block with this marching left-to-right, bottom-to-top procedure.

In the general scheme with $k \times k$ blocks, next use the matrices T_2 and T_3 to zero-out the block matrix T_{23} . (See below.)

$$T = \left[\begin{array}{ccc|cc} T_1 & 0 & T_{13} & \cdots & \cdots \\ & T_2 & T_{23} & & \\ & & T_3 & & \\ \hline & 0 & & \ddots & \\ & & & T_{k-1} & T_{k-1,k} \\ & & & & T_k \end{array} \right]$$

After that it is possible using blocks T_1 and T_3 to zero-out the block T_{13} . This is the general scheme for the entire triangular structure, moving down the super-diagonal $(j, j + 1)$ blocks and then up the (block) columns. In this way we can zero-out any value not in the square diagonal blocks pertaining to the eigenvalues $\lambda_1 \dots \lambda_k$. $\square$

Remark 4.2.1. If $A \in M_n(R)$ and $\sigma(A) \subset R$, then all operations can be carried out with real numbers.

Lemma 4.2.1. *Let $\mathcal{J} \subset M_n$ be a commuting family. Then there is a vector $x \in \mathbb{C}^n$ for which x is an eigenvector for every $A \in \mathcal{J}$.*

Proof. Let $W \subset \mathbb{C}^n$ be a subspace of minimal dimension that is invariant under $\mathcal{J}$. Since $\mathbb{C}^n$ is invariant, we know W exists. Since $\mathbb{C}^n$ is invariant, we know W exists. Suppose there is an $A \in \mathcal{J}$ for which there is a vector in W which is not an eigenvector of A . Define $W_0 = \{y \in W \mid Ay = \lambda y \text{ for some } \lambda\}$. That is, W_0 is a set of eigenvectors of A . Since W is invariant under A , it follows that $W_0 \neq \phi$. Also, by assumption. $W_0 \subsetneq W$. For any $x \in W_0$

$$ABx = (AB)x = B(Ax) = \lambda Bx$$

and so $Bx \in W_0$. It follows that W_0 is $\mathcal{J}$ invariant, and W_0 has lower positive dimension than W . $\square$

As a consequence we have the following result.

Theorem 4.2.3. *Let $\mathcal{J}$ be a commuting family in M_n . If $A \in \mathcal{J}$ is diagonalizable, then $\mathcal{J}$ is simultaneously diagonalizable.*

Proof. Since A is diagonalizable there are n linearly independent eigenvectors of A and if S is in M_n and consists of those eigenvectors the matrix $S^{-1}AS$ is diagonal. Since eigenvectors of A are the same as eigenvectors of $B \in \mathcal{J}$, it follows that $S^{-1}BS$ is also diagonal. Thus $S^{-1}\mathcal{J}S = \mathcal{D} :=$ all diagonal matrices. $\square$

Theorem 4.2.4. *Let $\mathcal{J} \subset M_n$ be a commuting family. There is a unitary matrix $U \in M_n$ such that U^*AU is upper triangular for each $A \in \mathcal{J}$.*

Proof. From the proof of Schur's Theorem we have that the eigenvectors chosen for U are the same for all $A \in \mathcal{J}$. This follows because after the first step we have reduced A and B to

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

respectively. Commutativity is preserved under simultaneous similarity and therefore A_{22} and B_{22} commutes. Therefore at the second step the same eigenvector can be selected for all $B_{22} \subset \mathcal{J}_2$, a commuting family. $\square$

Theorem 4.2.5. *If $A \in M_n(R)$, there is a real orthogonal matrix $Q \in M_n(R)$ such that*

$$(\star) \quad Q^T A Q = \begin{bmatrix} A_1 & & & \\ & A_2 & \star & \\ & 0 & \ddots & \\ & & & A_k \end{bmatrix} \quad 1 \leq k \leq n$$

where each A_i is a real 1×1 matrix or a real 2×2 matrix with a non real pair of complex conjugate eigenvalues.

Theorem 4.2.6. *Let $\mathcal{J} \subset M_n(R)$ be a commuting family. There is a real orthogonal matrix $Q \in M_n(R)$ for which $Q^T A Q$ has the form $(\star)$ for every $A \in \mathcal{J}$.*

Theorem 4.2.7. *Suppose $A, B \in M_n(\mathbb{C})$ have eigenvalues $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ respectively. If A and B commute then there is a permutation $i_1 \dots i_n$ of the integers $1, \dots, n$ for which the eigenvalues of $A + B$ are $\alpha_j + \beta_{i_j}$, $j = 1, \dots, n$. Thus $\sigma(A + B) \subset \sigma(A) + \sigma(B)$.*

Proof. Since $\mathcal{J} = \{A, B\}$ forms a commuting family we have that every eigenvector of A is an eigenvector of B , and conversely. Thus if $Ax = \alpha_j x$ we must have that $Bx = \beta_{i_j} x$. But how do we get the permutation? The answer is to simultaneously triangularize with $U \in M_n$. We have

$$U^* A U = T \quad \text{and} \quad U^* B U = R.$$

Since

$$U^*(A + B)U = T + R$$

we have that the eigenvalues pair up as described. $\square$

Note: We don't necessarily need A and B to commute. We need only the hypothesis that A and B are simultaneously diagonalizable.

If A and B do not commute little can be said of $\sigma(A + B)$. In particular $\sigma(A + B) \subsetneq \sigma(A) + \sigma(B)$. Indeed, by summing upper triangular and lower triangular matrices we can exhibit a range of possibilities. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, $\sigma(A + B) = \{-1, 1\}$ but $\sigma(A) = \sigma(B) = \{0\}$.

Corollary 4.2.1. *Suppose $A, B \in M_n$ are commuting matrices with eigenvalues*

$\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ respectively. If $\alpha_i \neq -\beta_j$ for all $1 \leq i, j \leq n$, then $A + B$ is nonsingular.

Note: Giving conditions for the nonsingularity of $A + B$ in terms of various conditions on A and B is a very different problem. It has many possible answers.

4.3 Exercises

1. Characterize all diagonal unitary matrices and all diagonal orthogonal matrices in M_n .
2. Let $T(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ where θ is any real. Then $T(\theta)$ is real orthogonal. Prove that if $U \in M_2(R)$ is real orthogonal, then U has the form $T(\theta)$ for some θ or

$$U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} T(\theta),$$

and conversely.

3. Let us define the $n \times n$ matrix P to be a w -permutation matrix if for every vector $x \in R_n$, the vector Px has the same components as x in value and number, though possibly permuted. Show that w -permutation matrices are permutation matrices.
4. In R_2 identify all Householder transformations that are rotations.
5. Given any unit vector w in the plane formed by e_i and e_j . Express the Householder matrix for this vector.
6. Show that the set of unitary matrices on $\mathbb{C}_n$ forms a subgroup of the subset of $GL(n, \mathbb{C})$.
7. In R_2 , prove that the product of two (Householder) reflections is a rotation. (Hint. If the reflection angles are θ_1 and θ_2 , then the rotation angle is $2(\theta_1 - \theta_2)$.)
8. Prove that the unitary matrices are closed the norm. That is, if the sequence $\{U_n\} \subset M_n(\mathbb{C})$ are all unitary and $\lim_{n \rightarrow \infty} U_n = U$ in the $\|\cdot\|_2$ norm, then U is also unitary.
9. Let $A \in M_n$ be invertible. Define $G = A^k, (A^{-1})^k, k = 1, 2, \dots$. Show that G is a subgroup of M_n . Here the group multiplication is matrix multiplication.

10. Prove that the unitary matrices are closed under pointwise convergence. That is, if the sequence $\{U_n\} \subset M_n(\mathbb{C})$ are all unitary and $\lim_{n \rightarrow \infty} U_n = U$ for each (ij) entry, then U is also unitary.
11. Suppose that $A \in M_n$ and that $AB = BA$ for all $B \in M_n$. Show that A is a multiple of the identity.
12. Prove that a unitary matrix U can be written as $V^{-1}W^{-1}VW$ for unitary V, W if and only if $\det U = 1$. (Bellman, 1970)
13. Prove that the only triangular unitary matrices are diagonal matrices.
14. We know that given any bounded sequence of numbers, there is a convergent subsequence. (Bolzano-Weierstrass Theorem). Show that the same is true for matrices for any given matrix norm. In particular, show that if U_n is any sequence of unitary matrices, then there is a convergent subsequence.
15. The *Hadamard Gate* (from quantum computing) is defined by the matrix $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Show that this transformation is a Householder matrix. What are its eigenvalues and eigenvectors?
16. Show that the unitary matrices do not form a subspace $M_n(\mathbb{C})$.
17. Prove that if a matrix $A \in M_n(\mathbb{C})$ preserves the orthonormality of one orthonormal basis, then it must be unitary.
18. Call the matrix A a **checkerboard matrix** if either

(I) $a_{ij} = 0$ if $i + j$ is even or (II) $a_{ij} = 0$ if $i + j$ is odd

We call the matrices of type I **even checkerboard** and type II **odd checkerboard**. Define $CH(n)$ to be all $n \times n$ invertible checkerboard matrices. The questions below all pertain to square matrices.

- (a) Show that if n is odd there are no invertible even checkerboard matrices.
- (b) Prove that every unitary matrix U has determinant with modulus one. (That is, $|\det U| = 1$.)
- (c) Prove that n is odd the product of odd checkerboard matrices is odd.

- (d) Prove that if n is even then the product of an even and an odd checkerboard matrix is odd, while the product of two even (or odd) checkerboard matrices is even.
- (e) Prove that if n is even then the inverse of any checkerboard matrix is a checkerboard matrix of the same type. However, in light of (a), it is only true that if n is an invertible odd checkerboard matrix, its inverse is odd.
- (f) Suppose that $n = 2m$. Characterize all the odd checkerboard invertible matrices.
- (g) Prove that $CH(n)$ is a subgroup of $GL(n)$.
- (h) Prove that the invertible odd checkerboard matrices of any size n forms a subgroup of $CH(n)$.

In connection with chess, checkerboard matrices give the type of chess board on which the maximum number of mutually non-attacking knights can be placed on the even ($i+j$ is even) or odd ($i+j$ is odd) positions.

- 19. Prove that every unitary matrix can be written as the product of a unitary diagonal matrix and another unitary matrix whose first column has nonnegative entries.

Chapter 5

Hermitian Theory

Hermitian matrices form one of the most useful classes of square matrices. They occur naturally in a variety of applications from the solution of partial differential equations to signal and image processing. Fortunately, they possess the most desirable of matrix properties and present the user with a relative ease of computation. There are several very powerful facts about Hermitian matrices that have found universal application. First the spectrum of Hermitian matrices is real. Second, Hermitian matrices have a complete set of orthogonal eigenvectors, which makes them diagonalizable. Third, these facts give a spectral representation for Hermitian matrices and a corresponding method to approximate them by matrices of less rank.

5.1 Diagonalizability of Hermitian Matrices

Let's begin by recalling the basic definition.

Definition 5.1.1. Let $A \in M_n(\mathbb{C})$. We say that A is *Hermitian* if $A = A^*$, where $A^* = \bar{A}^T$. A^* is called the adjoint of A . This, of course, is in conflict with the other definition of adjoint, which is given in terms of minors.

Recall the following facts and definitions about subspaces of $\mathbb{C}_n$:

- If U, V are subspaces of $\mathbb{C}_n$, we define the *direct sum* of U and V by $U \oplus V = \{u + v \mid u \in U, v \in V\}$.
- If U, V are subspaces of $\mathbb{C}_n$, we say U and V are orthogonal if $\langle u, v \rangle = 0$ for every $u \in U$ and $v \in V$. In this case we write $U \perp V$.

For example, a natural way to obtain orthogonal subspaces is from orthonormal bases. Suppose that $\{u_1, \dots, u_n\}$ is an orthonormal basis of $\mathbb{C}_n$. Let

the integers $\{1, \dots, n\}$ be divided into two (disjoint) subsets J_1 and J_2 . Now define

$$\begin{aligned} U_1 &= \mathfrak{S}\{u_i \mid i \in J_1\} \\ U_2 &= \mathfrak{S}\{u_i \mid i \in J_2\} \end{aligned}$$

Then U_1 and U_2 are orthogonal, i.e. $U_1 \perp U_2$, and

$$U_1 \oplus U_2 = \mathbb{C}_n$$

$$\langle Au, x \rangle = \langle u, Ax \rangle = \lambda \langle u, x \rangle = 0$$

Our main result is that Hermitian matrices are diagonalizable. To prove it, we reveal other interesting and important properties of Hermitian matrices. For example, consider the following.

Theorem 5.1.1. *Let $A \in M_n(\mathbb{C})$ be Hermitian. Then the spectrum of A , $\sigma(A)$, is real.*

Proof. Let $\lambda \in \sigma(A)$ with corresponding eigenvector $x \in \mathbb{C}_n$. Then

$$\begin{aligned} \langle Ax, x \rangle &= \langle x, Ax \rangle = \langle x, \lambda x \rangle = \bar{\lambda} \langle x, x \rangle \\ &\parallel \\ \langle \lambda x, x \rangle &= \lambda \langle x, x \rangle \\ &\parallel \\ &= \lambda \langle x, x \rangle. \end{aligned}$$

Since we know $\|x\|^2 = \langle x, x \rangle \neq 0$, it follows that $\lambda = \bar{\lambda}$, which is to say that λ is real. $\square$

Theorem 5.1.2. *Let $A \in M_n(\mathbb{C})$ be Hermitian and suppose that λ and μ are different eigenvalues with corresponding eigenvectors x and y . Then $x \perp y$ (i.e. $\langle x, y \rangle = 0$).*

Proof. We know $Ax = \lambda x$ and $Ay = \mu y$. Now compute

$$\begin{aligned} \langle Ax, y \rangle &= \langle x, A^*y \rangle = \langle x, Ay \rangle = \mu \langle x, y \rangle \\ &\parallel \\ &= \lambda \langle x, y \rangle. \end{aligned}$$

If $\langle x, y \rangle \neq 0$, the equality above yields a contradiction and the result is proved. $\square$

Remark 5.1.1. This result also follows from the previously proved result about the orthogonality of left and right eigenvectors pertaining to different eigenvalues.

Theorem 5.1.3. *Let $A \in M_n(\mathbb{C})$ be Hermitian, and let λ be an eigenvalue of A . Then the algebraic and geometric multiplicities of λ are equal. In symbols,*

$$m_a(\lambda) = m_g(\lambda).$$

Proof. We prove this result by reconsideration of our main result on triangularization of A by a similarity transformation. Let $x \in \mathbb{C}^n$ be an eigenvector of A pertaining to λ . So, $Ax = \lambda x$. Let $u_2, \dots, u_n \in \mathbb{C}^n$ be a set of vectors orthogonal to x , so that $\{x, u_2, \dots, u_n\}$ is a basis of $\mathbb{C}^n$. Indeed, it is an orthogonal basis, and by normalizing the vectors it becomes an orthonormal basis. We claim that $U_2 = \mathfrak{S}(u_2, \dots, u_n)$, the span of $\{u_2, \dots, u_n\}$ is invariant under A . To see this, suppose

$$u = \sum_{j=2}^n c_j u_j \in U_2$$

and

$$Au = v + ax$$

where $v \in U_2$ and $a \neq 0$. Then, on the one hand

$$\langle Au, x \rangle = \langle u, Ax \rangle = \lambda \langle u, x \rangle = 0$$

On the other hand

$$\langle Au, x \rangle = \langle v + ax, x \rangle = a \langle x, x \rangle = a \|x\|^2 \neq 0$$

This contradiction establishes that the span of $\{u_2, \dots, u_n\}$ is invariant under A .

Let $U \oplus V = \mathbb{C}^n$ be invariant subspaces with $U \perp V$. Suppose that U_B and V_B are orthonormal bases of the orthogonal subspaces U and V . Define the matrix $P \in M_n(\mathbb{C})$ by taking for its columns first the basis vectors U_B and then the basis vectors V_B . Let us write $P = [U_B, V_B]$ (with only a small abuse of notation). Then, since A is Hermitian,

$$\begin{aligned} B &= P^{-1}AP \\ &= \left[\begin{array}{c|c} A_u & 0 \\ \hline 0 & A_v \end{array} \right] \end{aligned}$$

The whole process can be carried out exactly $m_a(\lambda)$ times, each time generating a new orthogonal eigenvector pertaining to λ . This establishes that $m_g(\lambda) = m_a(\lambda)$. A formal induction could have been given. $\square$

Remark 5.1.2. Note how we applied orthogonality and invariance to force the triangular matrix of the previous result to become diagonal. This is what permitted the successive extraction of eigenvectors. Indeed, if for any eigenvector x the subspace of $\mathbb{C}_n$ orthogonal to x is invariant, we could have carried out the same steps as above.

We are now in a position to state our main result, whose proof is implicit in the three lemmas above.

Theorem 5.1.4. *Let $A \in M_n(\mathbb{C})$ be Hermitian. Then A is diagonalizable. The matrix P for which $P^{-1}AP$ is diagonal can be taken to be orthogonal. Finally, if $\{\lambda_1, \dots, \lambda_n\}$ and $\{u_1, \dots, u_n\}$ denote eigenvalues and pertaining orthonormal eigenvectors for A , then A admits the spectral representation $A = \sum_{j=1}^n \lambda_j u_j u_j^T$.*

Corollary 5.1.1. *Let $A \in M_n(\mathbb{C})$ be Hermitian.*

- (i) *A has n linearly independent and orthogonal eigenvectors.*
- (ii) *A is unitarily equivalent to a diagonal matrix.*
- (iii) *If $A, B \in M_n$ are unitarily equivalent, then A is Hermitian if and only if B is Hermitian.*

Note that in part (iii) above, the condition of unitary equivalence cannot be replaced by just similarity. (Why?)

Theorem 5.1.5. *If $A, B \in M_n(\mathbb{C})$ and $A \sim B$ with S as the similarity transformation matrix, $B = S^{-1}AS$. If $Ax = \lambda x$ and $y = S^{-1}x$, then $By = \lambda y$.*

If matrices are similar so also are their eigenstructures. It should establish the very closeness that similarity implies. Later as we consider decomposition theorems, we will see even more remarkable consequences.

Though we have as yet no method of determining the eigenvalues of a matrix beyond factoring the characteristic polynomial, it is instructive to see how their existence impacts the fundamental problem of solving $Ax = b$. Suppose that A is Hermitian with eigenvalues $\lambda_1, \dots, \lambda_n$, counted according to multiplicity and with orthonormal eigenvectors $\{u_1, \dots, u_n\}$. Consider

the following solution method for the system $Ax = b$. Since the span of the eigenvectors is C_n then

$$b = \sum_{i=1}^n b_i u_i$$

where as we know by the orthonormality of the vectors $\{u_1, \dots, u_n\}$ that $b_i = \langle b, u_i \rangle$. We can also write $x = \sum_{i=1}^n x_i u_i$. Then, the system becomes

$$\begin{aligned} Ax &= A \left(\sum_{i=1}^n x_i u_i \right) \\ &= \sum_{i=1}^n x_i \lambda_i u_i = \sum_{i=1}^n b_i u_i \end{aligned}$$

Therefore, the solution is

$$x_i = \frac{b_i}{\lambda_i}, \quad i = 1, \dots, n$$

Expanding the data vector b in the basis of eigenvectors yields a rapid method to find the solution to the system. Nonetheless, this is not the preferred method for solving linear systems when the coefficient matrix is Hermitian. Finding all the eigenvectors is usually costly, and other ways are available that are more efficient. We will discuss a few of them in the section and in later chapters.

Approximating Hermitian matrices

With the spectral representation available, we have a tool to approximate the matrix, keeping the “important” part and discarding the less important part. Suppose the eigenvalues are arranged in decending order $|\lambda_1| \geq \dots \geq |\lambda_n|$. Now approximate A by

$$A_k = \sum_{j=1}^k \lambda_j u_j u_j^T \tag{1}$$

This is an $n \times n$ matrix. The difference $A - A_k = \sum_{j=k+1}^n \lambda_j u_j u_j^T$. We can approximate the norm of the difference by

$$(A - A_k)x = \left(\sum_{j=k+1}^n \lambda_j u_j u_j^T \right) x = \sum_{j=k+1}^n \lambda_j x_j u_j$$

where $x = \sum_{j=1}^n x_j u_j$. Assume $\|x\| = 1$. By the Cauchy-Schwartz inequality

$$\begin{aligned} \|(A - A_k)x\|^2 &= \left\| \sum_{j=k+1}^n \lambda_j x_j u_j \right\|^2 \\ &\leq \sum_{j=k+1}^n |\lambda_j|^2 \end{aligned}$$

Therefore, $\|(A - A_k)\| \leq \left(\sum_{j=k+1}^n |\lambda_j|^2 \right)^{1/2}$. From this we can conclude that if the smaller eigenvalues are sufficiently small, the matrix can be accurately approximated by a matrix of lesser rank.

Example 5.1.1. The matrix

$$A = \begin{bmatrix} 0.5745 & -0.5005 & 0.1005 & 0.0000 \\ -0.5005 & 1.176 & -0.5756 & 0.1005 \\ 0.1005 & -0.5756 & 1.176 & -0.5005 \\ 0.0000 & 0.1005 & -0.5005 & 0.5745 \end{bmatrix}$$

has eigenvalues eigenvectors $\{2.004, 0.9877, 0.3219, 0.1872\}$ with pertaining eigenvectors

$$u_1 = \begin{bmatrix} 0.2740 \\ -0.6519 \\ 0.6519 \\ -0.2740 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0.4918 \\ -0.5080 \\ -0.5080 \\ 0.4918 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 0.6519 \\ 0.2740 \\ -0.2740 \\ -0.6519 \end{bmatrix}, \quad u_4 = \begin{bmatrix} 0.5080 \\ 0.4918 \\ 0.4918 \\ 0.5080 \end{bmatrix}$$

respectively. Neglecting the eigenvectors pertaining to the two smaller eigenvalues A is approximated according as 1 the formula above by

$$\begin{aligned} A_2 &= \sum_{j=1}^2 \lambda_j u_j u_j^T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T \\ &= 2.004 \begin{bmatrix} 0.274 \\ -0.6519 \\ 0.6519 \\ -0.274 \end{bmatrix} \begin{bmatrix} 0.274 \\ -0.6519 \\ 0.6519 \\ -0.274 \end{bmatrix}^T + 0.9877 \begin{bmatrix} 0.4918 \\ -0.508 \\ -0.508 \\ 0.4918 \end{bmatrix} \begin{bmatrix} 0.4918 \\ -0.508 \\ -0.508 \\ 0.4918 \end{bmatrix}^T \\ A_2 &= \begin{bmatrix} 0.3893 & -0.6047 & 0.1112 & 0.0884 \\ -0.6047 & 1.107 & -0.5968 & 0.1112 \\ 0.1112 & -0.5968 & 1.107 & -0.6047 \\ 0.0884 & 0.1112 & -0.6047 & 0.3893 \end{bmatrix} \end{aligned}$$

The difference

$$A - A_2 = \begin{bmatrix} 0.1852 & 0.1042 & -0.0107 & -0.0884 \\ 0.1042 & 0.069 & 0.0212 & -0.0107 \\ -0.0107 & 0.0212 & 0.069 & 0.1042 \\ -0.0884 & -0.0107 & 0.1042 & 0.1852 \end{bmatrix}$$

has 2-norm $\|A - A_2\|_2 = 0.3218$, while the 2-norm $\|A\|_2 = 2.004$. The relative error of approximation is $\frac{\|A - A_2\|_2}{\|A\|_2} = \frac{0.3218}{2.004} = 0.1606$.

To illustrate how this may be used, let us attempt to use A_2 to approximate the solution of $Ax = b$, where $b = [2.606, -4.087, 1.113, 0.3464]^T$. First of all the exact solution is $x = [2.223, -2.688, -0.1629, 0.9312]^T$. Since the matrix A_2 has rank two, it is not solvable for every vector b . We therefore project the vector b into the span of the range of A_2 , namely u_1 and u_2 . Thus

$$b_2 = \langle b, u_1 \rangle u_1 + \langle b, u_2 \rangle u_2 = [2.555, -4.118, 1.109, 0.3584]^T$$

Now solve $A_2 x_2 = b_2$, to obtain $x_2 = [2.023, -2.828, -0.2202, 0.9274]^T$. The 2-norm of the difference is $\|x - x_2\|_2 = 0.2508$. This error, though not extremely small, can be accounted for by the fact that the data vector b has sizable u_3 and u_4 components. That is $\|proj_{\mathfrak{S}(u_3, u_4)} b\| = 0.2508$.

5.2 Finding eigenvectors

Recall that a zero of a polynomial is called **simple** if its multiplicity is one. If the eigenvalues of $A \in M_n(\mathbb{C})$ are distinct and the largest, λ_n , in modulus is simple, then there is an iterative method to find it. Assume

- (1) $|\lambda_n| = \rho(A)$
- (2) $m_a(\lambda_n) = 1$.

Moreover, without loss of generality we assume eigenvalues to be ordered $|\lambda_1| \leq \dots \leq |\lambda_{n-1}| < |\lambda_n| = \rho_n$. Select $x^{(0)} \in \mathbb{C}^{(n)}$. Define

$$x^{(k+1)} = \frac{1}{\|x^{(k)}\|} Ax^{(k)}.$$

By scaling we can assume that $\lambda_n = 1$, and that $y^{(1)}, \dots, y^{(n)}$ are linearly independent eigenvectors of A . So $Ay^{(n)} = y^{(n)}$. We can write

$$x^{(0)} = c_1 y^{(1)} + \dots + c_n y^{(n)}.$$

Then, except for a scale factor (i.e. the factor $\|x^{(k)}\|^{-1}$)

$$x^{(k)} = c_1 \lambda_1^k y^{(1)} + \cdots + c_n \lambda_n^k y^{(n)}.$$

Since $|\lambda_j| < 1$, $j = 1, 2, \dots, n-1$, we have that $|\lambda_j^k| \rightarrow 0$ if $j = 1, 2, \dots, n-1$. Therefore, the limit of $x^{(k)}$ approaches a multiple of $y^{(n)}$.

This gives the following result.

Theorem 5.2.1. *Let $A \in M_n(\mathbb{C})$ have n distinct eigenvalues and assume the eigenvalue λ_n with modulus $\rho(A)$ is simple. If $x^{(0)}$ is not orthogonal to the eigenvector $y^{(n)}$ pertaining to λ_n , then the sequence of vectors defined by*

$$x^{(k+1)} = \frac{1}{\|x^{(k)}\|} Ax^{(k)}$$

converges to a multiple of $y^{(n)}$. This is called the Power Method.

The rate of convergence is controlled by $|\lambda_{n-1}|$. The closer to 1 this number is the slower the iterates converge. Also, if we know only that λ_n (for which $|\lambda_n| = \rho(A)$ is simple we can determine what it is by considering the Rayleigh quotient. Take

$$\rho_k = \frac{\langle Ax^{(k)}, x^{(k)} \rangle}{\langle x^{(k)}, x^{(k)} \rangle}.$$

Then $\lim_{k \rightarrow \infty} \rho_k = \lambda_n$. Thus the multiple of $y^{(n)}$ is indeed λ_n .

To find intermediate eigenvalues and eigenvectors we apply an adaptation of the power method called the *orthogonalization method*. However, in order to adapt the power method to determine λ_{n-1} , our underlying assumption is that is also simple and moreover $|\lambda_{n-2}| < |\lambda_{n-1}|$.

Assume $y^{(n)}$ and λ_n are known. Then we restart the iteration, taking the starting value

$$\hat{x}^{(0)} = x^{(0)} - \frac{\langle x^{(0)}, y^{(n)} \rangle}{\|y^{(n)}\|^2} y^{(n)}.$$

We know that the eigenvector $y^{(n-1)}$ pertaining to λ_{n-1} is orthogonal to $y^{(n)}$. Thus, in theory all of the iterates

$$\hat{x}^{(k+1)} = \frac{1}{\|\hat{x}^{(k)}\|} T \hat{x}^{(k)}$$

will remain orthogonal to $y^{(n)}$. Therefore,

$$\lim_{k \rightarrow \infty} \hat{x}^{(k)} = y^{(n-1)}.$$

—in theory. In practice, however, we must accept that $y^{(n)}$ has not been determined exactly. This means $\hat{x}^{(0)}$ has not been *purged* of all of $y^{(n)}$. By our previous reasoning, since λ_n is the dominant eigenvalue, the presence of $y^{(n)}$ will *creep* back into the iterates $\hat{x}^{(k)}$. To reduce the *contamination* it is best to *purify* the iterates $\hat{x}^{(k)}$ periodically by the reduction

$$(\star) \quad \hat{x}^{(k)} \longrightarrow \hat{x}^{(k)} - \frac{\langle \hat{x}^{(k)}, y^{(n)} \rangle}{\|y^{(n)}\|} y^{(n)}$$

before computing $\hat{x}^{(k+1)}$. The previous argument can be applied to prove that

$$(2) \quad \lim_{k \rightarrow \infty} x^{(k)} = y^{(n-1)}$$

$$(2) \quad \lim_{k \rightarrow \infty} \frac{\langle Ax^{(k)}, x^{(k)} \rangle}{\|x^{(k)}\|^2} = \lambda_{n-1}.$$

Additionally, even if we know $y^{(n)}$ exactly, round-off error would reinstate a $y^{(n)}$ component in our iterative computations. Thus the purification step above, $(\star)$, should be applied in *all* circumstances.

Finally, subsequent eigenvalues and eigenvectors may be determined by successive orthogonalizations. Again the eigenvalue simplicity and strict inequality is needed for convergence. Specifically, all eigenvectors can be determined if we assume that eigenvalues to be strictly ordered $|\lambda_1| < \dots < |\lambda_{n-1}| < |\lambda_n| = \rho_n$. For example, we begin the iterations to determine $y^{(n-j)}$ with

$$\hat{x}^{(0)} = x^{(0)} - \sum_{i=0}^{j-1} \frac{\langle x^{(0)}, y^{(n-i)} \rangle}{\|y^{(n-i)}\|^2} y^{(n-i)}.$$

Don't forget the re-orthogonalizations periodically throughout the iterative process.

What can be done to find intermediate eigenvalues and eigenvectors in the case A is not symmetric? The method above fails, but a variation of it works.

What must be done is to generate the left and right eigenvectors, $w^{(n)}$ and $y^{(n)}$, for A . Use the same process. To compute $y^{(n-1)}$ and $w^{(n-1)}$ we

orthogonalize thusly:

$$\begin{aligned}\hat{x}^{(0)} &= x^{(0)} - \frac{\langle x^{(0)}, w^{(n)} \rangle}{\|w^{(n)}\|^2} w^{(n)} \\ \hat{z}^{(0)} &= z^{(0)} - \frac{\langle z^{(0)}, y^{(n)} \rangle}{\|y^{(n)}\|^2} y^{(n)}\end{aligned}$$

where $z^{(0)}$ is the original starting value used to determine the left eigenvector $w^{(n)}$. Since we know that

$$\lim_{k \rightarrow \infty} x^{(k)} = \alpha y^{(n)}$$

it is easy to see that

$$\lim_{k \rightarrow \infty} Ax^{(k)} = \alpha \lambda_n y^{(n)}.$$

Therefore,

$$\lim_{k \rightarrow \infty} \frac{\langle Ax^{(k)}, x^{(k)} \rangle}{\langle x^{(k)}, x^{(k)} \rangle} = \lambda_n.$$

Example 5.2.1. Let T be the transformation of $R^2 \rightarrow R^2$ that rotates a vector by θ radians. Then it is clear that no matter what nonzero vector $x^{(0)}$ is selected the iterations $x^{(k+1)} = Tx^{(k)}$ will never converge. (Assume $\|x^{(0)}\| = 1$.) Now the matrix representation of T is

$$A_T = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad \text{rotates counterclockwise}$$

we have

$$p_{A_T}(\lambda) = \det \begin{bmatrix} \lambda - \cos \theta & \sin \theta \\ -\sin \theta & \lambda - \cos \theta \end{bmatrix} = (\lambda - \cos \theta)^2 + \sin^2 \theta.$$

The spectrum of A_T is therefore

$$\lambda = \cos \theta \pm i \sin \theta.$$

Notice that the eigenvalues are discrete, but there are two eigenvalues with modulus $\rho(A_T) = 1$. The above results therefore do not apply.

Example 5.2.2. Although the previous example is not based on a symmetric matrix, it certainly illustrates non convergence of the power iterations. The even simpler Householder matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ furnishes us with a symmetric matrix for which the iterations also do not converge. In this case, there are two eigenvalues with modulus equal to the spectral radius (± 1). With arbitrary starting vector $x^{(0)} = [a, b]^T$, it is obvious that the even iterations are $x^{(2i)} = [a, b]^T$ and the odd iterations are $x^{(2i-1)} = [a, -b]^T$.

Assuming again that the eigenvalues are distinct and even stronger, assuming that

$$|\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$$

we can apply the process above to extract all the eigenvalues (Rayleigh quotient) and the eigenvectors, one-by-one, when **A is symmetric**.

First of all, considering the matrix $A - \sigma I$ we can shift the eigenvalues to either the left or the right. Depending on the location of λ_n a sufficiently large $|\sigma|$ may be chosen so that $|\lambda_1 - \sigma| = \rho(A - \sigma I)$. The power method can be applied to determine $\lambda_1 - \sigma$ and hence λ_1 .

5.3 Positive definite matrices

Of the many important subclasses of Hermitian matrices, there is one class that stands out.

Definition 5.3.1. We say that $A \in M_n(\mathbb{C})$ is *positive definite* if $\langle Ax, x \rangle > 0$ for every nonzero $x \in C_n$. Similarly, we say that $A \in M_n(\mathbb{C})$ is *positive semidefinite* if $\langle Ax, x \rangle \geq 0$ for every nonzero $x \in C_n$.

It is easy to see that for positive definite matrices all of the results are true

Theorem 5.3.1. Let $A, B \in M_n(\mathbb{C})$. Then

1. If A is positive definite, then $\sigma(A) \subset \mathbb{R}_n^+$
2. If A is positive definite, then A is invertible.
3. B^*B is positive semidefinite.
4. If B is invertible then B^*B is positive definite.

5. If $B \in M_n(C)$ is positive semidefinite, then $\text{diag}(B)$ is nonnegative, and $\text{diag}(B)$ is strictly positive when B is positive definite.

The proofs are all routine. Of course, every diagonal matrix with non-negative entries is positive semidefinite.

Square roots

Given a real matrix $A \in M_n$. It is sometimes desired to determine a square root of A . By this we mean any matrix B for which $B^2 = A$. Moreover, if possible, it is desired that the square root be real. Our experience with numbers indicates that in order that a number have a positive square root, it must be positive. The analogue for matrices is the condition of being positive definite.

Theorem 5.3.2. *Let $A \in M_n$ be positive [semi-]definite. Then A has a real square root. Moreover, the square root can be taken to be positive [semi-]definite.*

Proof. We can write the diagonal matrix of the eigenvalues of A in the equation $A = P^{-1}DP$. Extract the positive square root of D as $D^{\frac{1}{2}} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_n^{1/2})$. Obviously $D^{\frac{1}{2}}D^{\frac{1}{2}} = D$. Now define $A^{\frac{1}{2}}$ by $A^{\frac{1}{2}} = P^{-1}D^{\frac{1}{2}}P$. This matrix is real. It is simple to check that $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$, and that this particular square root is positive definite. $\square$

Clearly any real diagonalizable matrix with nonnegative eigenvalues has a real square root as well. However, beyond that conditions for determining existence let alone determination of square roots take us into a very specialized subject.

5.4 Singular Value Decomposition

Definition 5.4.1. For any $A \in M_{mn}$, the $n \times n$ Hermitian matrix A^*A is positive semi-definite. Denoting its eigenvalues by λ_j we call the values $\sqrt{\lambda_j}$ the **singular values** of A .

Because $r(A^*A) \leq \min(r(A^*), r(A)) \leq \min(m, n)$ there are at most $\min(m, n)$ nonzero singular values.

Lemma 5.4.1. *Let $A \in M_{mn}$. There is an orthonormal basis $\{u_1, \dots, u_n\}$ of C_n such that $\{Au_1, \dots, Au_n\}$ is orthogonal.*

Proof. Proof. Consider the $n \times n$ Hermitian matrix A^*A , and denote an orthonormal basis of its eigenvectors by $\{u_1, \dots, u_n\}$. Then it is easy to see that $\{Au_1, \dots, Au_n\}$ is an orthogonal set. For $\langle Au_j, Au_k \rangle = \langle A^*Au_j, u_k \rangle = \lambda_j \langle u_j, u_k \rangle = 0$. $\square$

Lemma 5.4.2. *Lemma 2 Let $A \in M_{mn}$ and an orthonormal basis $\{u_1, \dots, u_n\}$ of C_n . Define*

$$v_j = \begin{cases} \frac{1}{\|Au_j\|} Au_j & \text{if } \|Au_j\| \neq 0 \\ 0 & \text{if } \|Au_j\| = 0 \end{cases}$$

Let $S = \text{diag}(\|Au_1\|, \dots, \|Au_n\|)$, the $n \times n$ matrix U having rows given by the basis $\{u_1, \dots, u_n\}$ and $\hat{V}$ the $m \times n$ matrix given by the columns $\{v_1, \dots, v_n\}$. Then $A = \hat{V}SU$.

Proof. Proof. Consider

$$\begin{aligned} \hat{V}SUu_j &= \begin{bmatrix} v_1 & v_2 & v_n \\ \downarrow & \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \|Au_1\| & 0 & \cdots & 0 \\ 0 & \|Au_2\| & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \|Au_n\| \end{bmatrix} \begin{bmatrix} u_1 \longrightarrow \\ u_2 \longrightarrow \\ \vdots \\ u_n \longrightarrow \end{bmatrix} u_j \\ &= \begin{bmatrix} v_1 & v_2 & v_n \\ \downarrow & \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \|Au_1\| & 0 & \cdots & 0 \\ 0 & \|Au_2\| & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \|Au_n\| \end{bmatrix} e_j \\ &= \begin{bmatrix} v_1 & v_2 & v_n \\ \downarrow & \downarrow & \downarrow \end{bmatrix} \|Au_j\| e_j = \|Au_j\| v_j = Au_j \end{aligned}$$

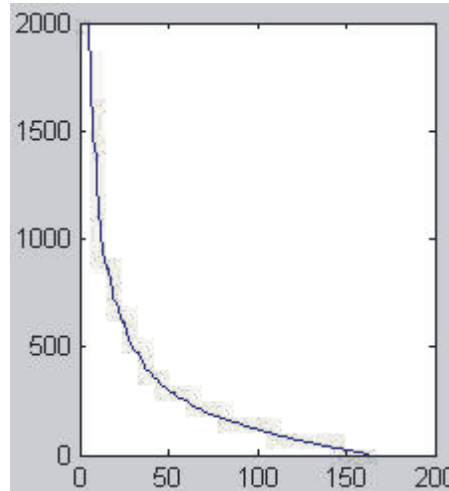
Thus both $\hat{V}SU$ and A have the same action on a basis. Therefore they are equal. $\square$

It is easy to see that $\|Au_j\| = \sqrt{\lambda_j}$, that is the singular values. While $A = \hat{V}SU$ could be called the singular value decomposition (SVD), what is usually offered at the SVD is small modification of it. Redefine the matrix U so that the first r columns pertain to the nonzero singular values. Define D to be the $m \times n$ matrix consisting of the non zero singular values in the

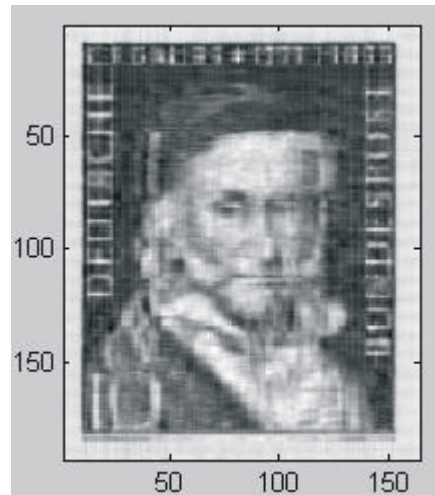
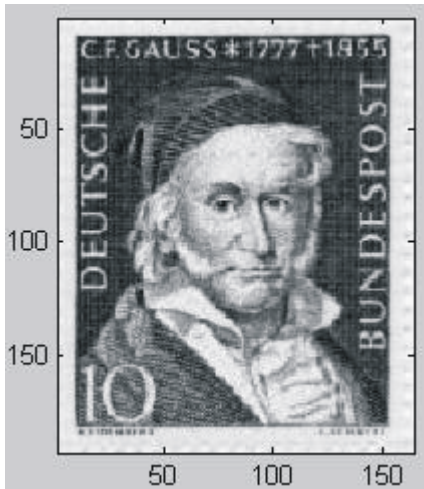
d_{jj} positions, and filled in with zeros else where. Define the matrix V to be the first r of the columns of $\hat{V}$ and if $r < m$ construct an additional $m - r$ orthonormal columns so that V is an orthonormal basis of $\mathbf{C}_m$. The resulting product VDU , called the **singular value decomposition** of A , is equal to A , and moreover it follows that V is $m \times m$, D is $m \times n$, and U is $n \times n$. This gives the following theorem

Theorem 5.4.1. *Let $A \in M_{mn}$. Then there is an $m \times m$ orthogonal matrix V , an $n \times n$ orthogonal matrix U , and an $m \times n$ matrix D with only diagonal entries such that $A = VDU$. The diagonal entries of D are the singular values of A and the rows of U are the eigenvectors of A^*A .*

Example 5.4.1. The singular value decomposition can be used for image compression. Here is the idea. Consider all the eigenvalues of A^*A and order them greatest to least. Zero the matrix S for all eigenvalues less than some threshold. Then in the reconstruction and transmission of the matrix, it is not necessary to include the vectors pertaining to these eigenvalues. In the example below, we have considered a 164×193 pixel image of C. F. Gauss (1777-1855) on the postage stamp issued by Germany on Feb. 23, 1955, to commemorate the centenary of death. Therefore its spectrum has 164 eigenvalues. The eigenvalues range from 26,603.0 to 1.895. A plot of the eigenvalues shown below. Now compress the image, retaining only a fraction of the eigenvalues by effectively zeroing the smaller eigenvalues.



Note that the original image of the stamp has been enlarged and resampled for more accurate comparisons. This image (stored at 221 dpi) is displayed at effective 80 dpi with the enlargement. Below we show two plots where we have retained respectively 30% and 10% of the eigenvalues. There is an apparent drastic decline in the image quality at roughly 10:1 compression. In this image all eigenvalues smaller than $\lambda = 860$ have been zeroed.



Using 48 of 164 eigenvalues Using 16 of 164 eigenvalues

5.5 Exercises

1. Prove Theorem 5.3.1 (i).
2. Prove Theorem 5.3.1 (ii).
3. Suppose that $A, B \in M_n(\mathbb{C})$ are Hermitian. We will say $A \succcurlyeq 0$ if A is non-negative definite. Also, we say $A \succcurlyeq B$ if $A - B \succcurlyeq 0$. Is “ $\succcurlyeq$ ” an equivalence relation? If $A \succcurlyeq B$ and $B \succcurlyeq C$ prove or disprove that $A \succcurlyeq C$.
4. Describe all Hermitian matrices of rank one.
5. Suppose that $A, B \in M_n(\mathbb{C})$ are Hermitian and positive definite. Find necessary and sufficient conditions for AB to be Hermitian and also positive definite.
6. For any matrix $A \in M_n(\mathbb{C})$ with eigenvalues λ_i , $i = 1, \dots, n$. Prove that $\sum_{i=1}^n |\lambda_i|^2 = \sum_{i,j=1}^n |a_{ij}|^2$.

Chapter 6

Normal Matrices

Normal matrices are matrices that include Hermitian matrices and enjoy several of the same properties as Hermitian matrices. Indeed, while we proved that Hermitian matrices are unitarily diagonalizable, we did not establish any converse. That is, if a matrix is unitarily diagonalizable, then does it have any special property involving for example its spectrum or its adjoint? As we shall see normal matrices are unitarily diagonalizable.

6.1 Introduction to Normal matrices

Definition 6.1.1. A matrix $A \in M_n$ is called *normal* if $A^*A = AA^*$.

Proposition 6.1.1. $A \in M_n$ is normal if and only if every matrix unitarily equivalent to A is normal.

Proof. Suppose A is normal and $B = U^*AU$, where U is unitary. Then $B^*B = U^*A^*AU = U^*AA^*U = U^*AUU^*A^*U = BB^*$. If U^*AU is normal then it is easy to see that $U^*AA^*U = U^*A^*AU$. Multiply this equation on the right by U^* and on the left by U to obtain $AA^* = A^*A$. $\square$

Examples.

- (1) Unitary matrices are normal ($U^*U = I = UU^*$).
- (2) Hermitian matrices are normal ($AA^* = A^2 = A^*A$).
- (3) If $A^* = -A$, we have $A^*A = AA^* = -A^2$. Hence matrices for which $A^* = -A$, called *skew-Hermitian*, are normal.

Example 6.1.1. Consider the arbitrary matrix $N \in M_2(R)$, written as $N = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If we suppose that N is normal then

$$\begin{aligned} N^*N &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}^T \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix} \\ NN^* &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} \end{aligned}$$

From this we conclude that $b^2 = c^2$, or $b = \pm c$. Consider the cases in turn.

(i) If $c = b$, then N is Hermitian and thus normal.

(ii) If $c = -b \neq 0$, then $(N^*N)_{12} = ab + cd = b(a - d)$. On the other hand $(NN^*)_{12} = ac + bd = (d - a)b$. For $b(a - d) = (d - a)b$, we must have $a = d$. This gives that real 2×2 normal matrices are either symmetric or have the form

$$N = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

Note this form includes both rotations and skew-symmetric matrices.

Recall the definition of a unitarily diagonalizable matrix: A matrix $A \in M_n$ is called *unitarily diagonalizable* if there is a unitary matrix U for which U^*AU is diagonal. A simple consequence of this is that if $U^*AU = D$ (where $D = \text{diagonal}$ and $U = \text{unitary}$), then

$$AU = UD$$

and hence A has n orthonormal eigenvectors. This is just a part of the *spectral theorem for normal matrices*.

Theorem 6.1.1 (Spectral theorem for normal matrices). *If $A \in M_n$ has eigenvalues $\lambda_1 \dots \lambda_n$, counted according to multiplicity, the following statements are equivalent.*

- (a) A is normal.
- (b) A is unitarily diagonalizable.
- (c) $\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 = \sum_{j=1}^n |\lambda_j|^2$.
- (d) There is an orthonormal set of n eigenvectors of A .

Proof. (a) $\Rightarrow$ (b). If A is normal, then AA^* is Hermitian and therefore unitarily diagonalizable. Thus $U^*A^*AU = D = U^*AA^*U$. Also, $A, A^*, A^*A = AA^*$ form a commuting family. This implies that eigenvectors of A^*A are also eigenvectors of A . Since A^*A has a complete orthonormal set we know that U^*AU is also diagonal. It is easy to see also that (b) $\Rightarrow$ (a). We also note that (a) $\Rightarrow$ (d).

(b) $\Rightarrow$ (c). Suppose $U^*AU = D$. Then $U^*A^*U = D^*$ and $U^*A^*AU = D^*D$. By Corollary 3.5.2 similarly preserves the trace. We know trace of A^*A is $\text{tr}(A^*A) = \sum_{j=1}^n \sum_{k=1}^n a_{jk}^* a_{kj} = \sum_{j=1}^n \sum_{k=1}^n \bar{a}_{kj} a_{kj} = \sum_{j=1}^n \sum_{k=1}^n |a_{kj}|^2$. Since the trace of D^*D is $\sum |\lambda_j|^2$, the result follows.

(c) $\Rightarrow$ (b). We know that A is unitarily equivalent to a upper triangular matrix T . We also know that if $A \sim B$ are unitarily equivalent $\sum_{ij} |a_{ij}|^2 = \sum_{ij} |b_{ij}|^2$. Application of this equality to the upper triangular matrix T yields

$$\sum_{i,j} |a_{ij}|^2 = \sum |\lambda_j|^2 + \sum_{j>i} |t_{ij}|^2 = \sum |\lambda_j|^2.$$

Thus $t_{ij} = 0$ for $j > i$. Thus A is uniformly diagonalizable.

(d) $\Rightarrow$ (b). Trivial. $\square$

Corollary 6.1.1. *Let $A \in M_n$ and A is normal. If U is unitary and if U^*AU is upper triangular then U^*AU is diagonal.*

Theorem 6.1.2. *Let $N \in M_n(\mathbb{R})$. Then N is normal if and only if there is a real orthogonal matrix $Q \in M_n(\mathbb{R})$ such that*

$$Q^T N Q = \begin{bmatrix} A_1 & & & \\ & A_2 & & \bigcirc \\ & & \ddots & \\ & \bigcirc & & A_n \end{bmatrix} \quad (1)$$

where A_i is 1×1 (real) or A_i is 2×2 (real) of the form

$$A_i = \begin{bmatrix} \alpha_i & \beta_j \\ -\beta_j & \alpha_i \end{bmatrix}.$$

Proof. First of all, any matrix A of the form given by (1) is normal, and therefore so also is any matrix unitarily similar (real orthogonally similar in this case) to it.

To prove the converse we assume that $N \in M_n(R)$ is normal. We know that N is unitarily diagonalizable. That is, there is a unitary matrix U such that $U^*NU = D$, the diagonal matrix of its eigenvalues. Because N is real, all complex eigenvalues occur in complex conjugate pairs. Arrange them as successive diagonal entries in D . If λ is a real eigenvalue, we can assume without loss of generality that the corresponding eigenvector is real. For complex eigenvalues, the corresponding eigenvectors also occur in conjugate pairs. Thus if $\alpha + i\beta$ is an eigenvalue of N with corresponding eigenvector written in real and complex parts $u = u_r + iu_s$. Since N is real we have that $\alpha - i\beta$ is also an eigenvalue of N with corresponding eigenvector $\bar{u} = u_r - iu_s$. By the fact that N is unitarily diagonalizable, these vectors are orthogonal. This means $\langle u_r, u_s \rangle = 0$.

Replace the eigenvectors $u_r \pm iu_s$ by the real and imaginary parts in U . This gives the matrix Q . Now compute Q^TNQ . It is easy to see that compute $Nu_r = \alpha u_r - \beta v_s$ and $Nu_s = \alpha u_s + \beta v_r$. When the first of these vectors $(\alpha u_r - \beta v_s)$ is multiplied by Q^T we obtain the vector $[0, \dots, \alpha, -\beta, 0, \dots, 0]^T$. Multiplication by the second gives the vector $[0, \dots, \beta, \alpha, 0, \dots, 0]^T$. In this way the components $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ arise. $\square$

Corollary 6.1.2. (a) $A \in M_n$ is symmetric if and only if (1) holds with all blocks 1×1 (and real).

(b) $AA^T = I$ if and only if (1) has the form

$$\begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_p & & \\ & & & A_1 & * \\ * & & & & \ddots & \\ & & & & & A_k \end{bmatrix}$$

where $\lambda_j = \pm 1$ and $A_j = \begin{bmatrix} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{bmatrix}$ $\theta_j \in R$.

6.2 Exercises

1. If A and B commute and if A is normal, then A^* and B commute.

Chapter 7

Factorization Theorems

This chapter highlights a few of the many factorization theorems for matrices. While some factorization results are relatively direct, others are iterative. While some factorization results serve to simplify the solution to linear systems, others are concerned with revealing the matrix eigenvalues. We consider both types of results here.

7.1 The PLU Decomposition

The PLU decomposition (or factorization) To achieve LU factorization we require a modified notion of the row reduced echelon form.

Definition 7.1.1. The *modified row echelon form* of a matrix is that form which satisfies all the conditions of the modified row reduced echelon form except that we do not require zeros to be above leading ones, and moreover we do not require leading ones, just nonzero entries.

For example the matrices below are in row echelon form.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & -7 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Most of the factorizations $A \in M_n(\mathbb{C})$ studied so far require one essential ingredient, namely the eigenvectors of A . While it was not emphasized when we studied Gaussian elimination, there is a LU-type factorization there. Assume for the moment that the only operations needed to carry A to its

modified row echelon form are those that add a multiple of one row to another. The *modified row echelon form* of a matrix is that form which satisfies all the conditions of the modified row reduced echelon form except that we do not require zeros to be above leading ones, and moreover we do not require leading ones, just nonzero entries. Naturally it is easy to make the leading nonzero entries into leading ones by the multiplication by an appropriate identity matrix. That is not the point here. What we want to observe is that in this case the reduction is accomplished by the left multiplication of A by a sequence of lower triangular matrices of the form.

$$L = \begin{bmatrix} 1 & & & \\ 0 & 1 & & 0 \\ \vdots & 0 & 1 & \\ & c & & \ddots \\ 0 & \dots & & & 1 \end{bmatrix}$$

Since we pivot at the $(1, 1)$ -entry first, we eliminate all the entries in the first column below the first row. The product of all the matrices L to accomplish this has the form

$$L_1 = \begin{bmatrix} 1 & & & \\ c_{21} & 1 & & 0 \\ c_{31} & 0 & 1 & \\ \vdots & & & \ddots \\ c_{n1} & 0 & \dots & & 1 \end{bmatrix}$$

where $c_{k1} = -\frac{a_{k1}}{a_{11}}$. Thus, with the notation that $A = A_1$ has entries $a_{ij}^{(1)}$ this first phase of the reduction renders the matrix A_2 with entries $a_{ij}^{(2)}$

$$A_2 = L_1 A_1 = \begin{bmatrix} a_{11}^{(2)} & \dots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \dots & \vdots \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \\ \vdots & & & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \dots & & a_{nn}^{(2)} \end{bmatrix}$$

Since we have assumed that no row interchanges are necessary to carry out the reduction we know that $a_{22}^{(2)} \neq 0$. The next part of the reduction process is the elimination of the elements in the second column below the second

row, i.e. $a_{32}^{(2)} \rightarrow 0, \dots a_{n2}^{(2)} \rightarrow 0$. Correspondingly, this can be achieved by a matrix of the form

$$L_2 = \begin{bmatrix} 1 & & & \\ 0 & 1 & & 0 \\ 0 & c_{22} & 1 & \\ \vdots & \vdots & & \ddots \\ 0 & c_{n2} & \dots & 1 \end{bmatrix}$$

(What are the values c_{k2} ?) The result is the matrix A_3 given by

$$A_3 = L_2 A_2 = L_2 L_1 A_1 = \begin{bmatrix} a_{11}^{(3)} & \dots & a_{1n}^{(3)} \\ 0 & a_{22}^{(3)} & \dots & \vdots \\ 0 & 0 & a_{33}^{(3)} & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{3n}^{(3)} & a_{nn}^{(3)} \end{bmatrix}$$

Proceeding in this way through all the rows (columns) there results

$$A_n = L_{n-1} A_{n-1} = L_{n-1} \dots L_2 L_1 A_1 = \begin{bmatrix} a_{11}^{(3)} & \dots & a_{1n}^{(3)} \\ 0 & a_{22}^{(3)} & \dots & \vdots \\ 0 & 0 & a_{33}^{(3)} & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_{nn}^{(3)} \end{bmatrix}$$

The right side of the equation above is an upper triangular matrix. Denote it by U . Since each of the matrices L_i , $i = 1, \dots, n-1$ is invertible we can write

$$A = L_1^{-1} \dots L_{n-1}^{-1} U$$

The lemma below is useful in this.

Lemma 7.1.1. *Suppose the lower triangular matrix $L \in M_n(\mathbb{C})$ has the*

form

$$L = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & & 0 \\ & & 1 & & & \\ & & 0 & 1 & & \\ \vdots & & \vdots & c_{k+1,k} & \ddots & \\ & & \vdots & & \ddots & \\ 0 & \cdots & 0 & c_{nk} & & 1 \end{bmatrix} \leftarrow k^{th} \text{ row}$$

Then L is invertible with inverse given by

$$L^{-1} = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & & 0 \\ & & 1 & & & \\ & & 0 & 1 & & \\ \vdots & & \vdots & -c_{k+1,k} & \ddots & \\ & & \vdots & & \ddots & \\ 0 & \cdots & 0 & -c_{nk} & & 1 \end{bmatrix} \leftarrow k^{th} \text{ row}$$

Proof. Trivial

Lemma 7.1.2. Suppose $L_1, L_2, \dots, L_{n-1}$ are the matrices given above. Then

the matrix $L = L_1^{-1} \cdots L_{n-1}^{-1}$ has the form

$$L = \begin{bmatrix} 1 & & & & & \\ -c_{21} & 1 & & & & 0 \\ -c_{31} & -c_{32} & 1 & & & \\ & & & 1 & & \\ \vdots & \vdots & \vdots & -c_{k+1,k} & \ddots & \\ & & & \vdots & \ddots & \\ -c_{n1} & -c_{n2} & \cdots & -c_{nk} & \cdots & 1 \end{bmatrix}$$

Proof. Trivial.

Applying these lemmas to the present situation we can say that when no row interchanges are needed we can factor and matrix $A \in M_n(\mathbb{C})$ as $A = LU$, where L is lower triangular and U is upper triangular. When row

interchanges are needed and we let P be the permutation matrix that creates these row interchanges then the LU-factorization above can be carried out for the matrix PA . Thus $PA = LU$, where L is lower triangular and U is upper triangular. We call this the **PLU** factorization. Let us summarize this in the following theorem.

Theorem 7.1.1. *Let $A \in M_n(\mathbb{C})$. Then there is a permutation matrix $P \in M_n(\mathbb{C})$ and lower L and upper U triangular matrices ($\in M_n(\mathbb{C})$), such that $PA = LU$. Moreover, L can be taken to have ones on its diagonal. That is, $\ell_{ii} = 1, \quad i = 1, \dots, n$.*

By applying the result above to A^T it is easy to see that the matrix U can be taken to have the ones in its diagonal. The result is stated as a corollary.

Corollary 7.1.1. *Let $A \in M_n(\mathbb{C})$. Then there is a permutation matrix $P \in M_n(\mathbb{C})$ and lower and upper triangular matrices ($\in M_n(\mathbb{C})$) respectively, such that $PA = LU$. Moreover, U can be taken to have ones on its diagonal ($u_{ii} = 1, \quad i = 1, \dots, n$).*

The PLU decomposition can be put in service to solving the system $Ax = b$ as follows. Assume that $A \in M_n(\mathbb{C})$ is invertible. Determine the permutation matrix P in order that $PA = LU$, where L is lower triangular and U is upper triangular. Thus, we have

$$\begin{aligned} Ax &= b \\ PAx &= Pb \\ LUx &= Pb \end{aligned}$$

Solve the systems

$$\begin{aligned} Ly &= Pb \\ Ux &= y \end{aligned}$$

Then $LUx = Ly = Pb$. Hence x is a solution to the system. The advantages of this formulation over the direct Gaussian elimination is that the systems $Ly = Pb$ and $Ux = y$ are triangular and hence are easy to solve. For example for the first of the systems, $Ly = Pb$, let the vector $Pb = [\hat{b}_1, \dots, \hat{b}_n]^T$. Then it is easy to see that “back substitution” (aka “forward substitution”)

can be used to determine y . That is, we have the recursive relations

$$\begin{aligned} y_1 &= \frac{\hat{b}_1}{l_{11}} \\ y_2 &= \frac{\hat{b}_2 - l_{21}y_1}{l_{22}} \\ &\vdots \\ y_n &= \left(\hat{b}_n - \sum_{m=1}^{n-1} l_{nm}y_m \right) l_{nn}^{-1} \end{aligned}$$

A similar formula applies to solve $Ux = y$. In this case we solve first for $x_n = y_n/u_{nn}$. The general formula is recursive with x_k being determined after $x_{k+1}, \dots, x_n$ are determined using the formula

$$x_k = \left(y_k - \sum_{m=k+1}^n u_{km}y_m \right) u_{kk}^{-1}$$

In practice the step of determining and then multiplying by the permutation matrix is not actually carried out. Rather, an index array is generated, while the elimination step is accomplished that effectively interchanges a “pointer” to the row interchanges. This saves considerable time in solving potentially very large systems.

More general and instructive methods are available for accomplishing this LU factorization. Also, conditions are available for when no (nontrivial) permutation is required. We need the following lemma.

Lemma 7.1.3. *Let $A \in M_n(\mathbb{C})$ have the LU factorization $A = LU$, where L is lower triangular and U is upper triangular. For any partition of the matrix of the form*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

there are corresponding decompositions of the matrices L and U

$$L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

where the L_{ii} and the U_{ii} are lower and upper triangular respectively. Moreover, we have

$$\begin{aligned} A_{11} &= L_{11}U_{11} \\ A_{21} &= L_{21}U_{11} \\ A_{12} &= L_{12}U_{22} \\ A_{22} &= L_{21}U_{12} + L_{22}U_{22} \end{aligned}$$

Thus $L_{11}U_{11}$ is a LU factorization of A_{11} .

With this lemma we can establish that almost every matrix can have a LU factorization.

Definition 7.1.2. Let $A \in M_n(\mathbb{C})$ and suppose that $1 \leq j \leq n$. The expression $\det(A\{1, \dots, j\})$ means the determinant of the upper left $j \times j$ submatrix of A . These quantities for $j = 1, \dots, n$ are called the **principal determinants of A** .

Theorem 7.1.2. Let $A \in M_n(\mathbb{C})$ and suppose that A has rank k . If

$$\det(A\{1, \dots, j\}) \neq 0 \quad \text{for } j = 1, \dots, k \quad (1)$$

then A has a LU factorization $A = LU$, where L is lower triangular and U is upper triangular. Moreover, the factorization may be taken so that either L or U is nonsingular. In the case $k = n$ both L and U will be nonsingular.

Proof. We carry out this LU factorization as a direct calculation in comparison to the Gaussian elimination method above. Let us propose to solve the equation $LU = A$ expressed as

$$\begin{bmatrix} l_{11} & & & & \\ l_{21} & l_{22} & & & 0 \\ l_{31} & l_{32} & l_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & \ddots & \\ l_{n1} & l_{n2} & \cdots & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ & u_{22} & u_{23} & \cdots & u_{2n} \\ & & u_{33} & & \\ & 0 & & \ddots & \vdots \\ & & & & \ddots & \\ & & & & & u_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & \ddots & \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}$$

It is easy to see that $l_{11}u_{11} = a_{11}$. We can take, for example $l_{11} = 1$ and solve for u_{11} . The determinant condition assures us that $u_{11} \neq 0$. Next solve for the $(2, 1)$ -entry. We have $l_{21}u_{11} = a_{21}$. Since $u_{11} \neq 0$, solve for l_{21} . For the $(1, 2)$ -entry we have $l_{11}u_{12} = a_{12}$, which can be solved for u_{12} since $l_{11} \neq 0$. Finally, for the $(2, 2)$ -entry, $l_{12}u_{12} + l_{22}u_{22} = a_{22}$ is an equation with two unknowns. Assign $l_{22} = 1$ and solve for u_{22} . What is important to note is that the process carried out this way gives the factorization of the upper left 2×2 submatrix of A . Thus

$$\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Since $\det \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right) \neq 0$, it follows that $\det \left(\begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} \right) \neq 0$ and we know that $\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$ is nonsingular as the diagonal elements are ones. Continue the factorization process through the $k \times k$ upper left submatrix of A .

Now consider the blocked matrix form form A

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $k \times k$ and has rank k . Thus we know that the rows of the lower $(n - k) \times n$ matrix above, that is $\begin{bmatrix} A_{21} & A_{22} \end{bmatrix}$ can be written as a unique linear combination of the rows of the upper $k \times n$ matrix $\begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$. Thus

$$\begin{bmatrix} A_{21} & A_{22} \end{bmatrix} = C \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

for some $(n - k) \times k$ matrix C . Of course this means: $A_{21} = C A_{11}$ and $A_{22} = C A_{12}$. We consider the factorization

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

where the blocks L_{11} and U_{11} have just been determined. From the equations in the lemma above we solve to get $U_{12} = L_{11}^{-1} A_{12}$ and $L_{21} =$

$A_{12}U_{11}^{-1}$. Then

$$\begin{aligned}
 A_{22} &= L_{21}U_{12} + L_{22}U_{22} \\
 &= A_{12}U_{11}^{-1}L_{11}^{-1}A_{12} + L_{22}U_{22} \\
 &= A_{12}A_{11}^{-1}A_{12} + L_{22}U_{22} \\
 &= C A_{11}A_{11}^{-1}A_{12} + L_{22}U_{22} \\
 &= C A_{12} + L_{22}U_{22} \\
 &= A_{22} + L_{22}U_{22}
 \end{aligned}$$

Thus we solve $L_{22}U_{22} = 0$. Obviously, we can take for L_{22} any nonsingular matrix we wish and solve for U_{22} or conversely.

□

7.2 **LR** factorization

While the PLU factorization is useful for solving systems, the LR factorization can be used to determine eigenvalues. .

Let $A \in M_n$ be given. Then

$$A = A_1 = L_1 R_1.$$

Then

$$\begin{aligned}
 L_1^{-1}A_1L_1 &= R_1L_1 \equiv A_2 \\
 A_2 &= L_2R_2 \\
 L_2^{-1}A_2L_2 &= R_2L_2 \equiv A_3.
 \end{aligned}$$

Continue in this fashion to obtain

$$L_k^{-1}A_kL_k = R_kL_k \equiv A_{k+1} \quad (\star)$$

We define

$$\begin{aligned}
 P_k &= L_1L_2 \dots L_k \\
 Q_k &= R_k \dots R_2R_1.
 \end{aligned}$$

Then

$$P_k A_{k+1} = A_1 P_k$$

for

$$\begin{aligned}
 A_{k+1} &= L_k^{-1} A_k L_k \\
 &= L_k^{-1} L_{k-1}^{-1} A_{k-1} L_{k-1} L_k \\
 &\vdots \\
 &= P_k^{-1} A_1 P_k
 \end{aligned}$$

or

$$P_k A_{k+1} = A_1 P_k.$$

Hence

$$\begin{aligned}
 P_k Q_k &= P_{k-1} A_k Q_{k-1} \\
 &= A_1 P_{k-1} Q_{k-1} \\
 &= A_1 P_{k-2} A_{k-1} Q_{k-2} \\
 &= A_1^2 P_{k-2} Q_{k-2} \\
 &\vdots \\
 &= A_1^k.
 \end{aligned}$$

Theorem 7.2.1 (Rutishauser). *Let $A \in M_n$ be given. Assume the eigenvalues of A satisfy*

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0.$$

Then $A \sim \Lambda = \text{diag}(\lambda_1 \dots \lambda_n)$. Assume $A = S\Lambda S^{-1}$, and

$$Y \equiv S^{-1} = L_y R_y \quad X = S = L_x R_x$$

where L_y and L_x are lower unit triangular matrices and R_y and R_x are upper triangular. Then A_k defined by $(\star)$ satisfy the result $\lim A_k$ is upper triangular.

Proof. (Wilkinson) We have

$$\begin{aligned}
 A_1^k &= X \Lambda^k Y \\
 &= X \Lambda^k L_y R_y \\
 &= X \Lambda^k L_y \Lambda^{-k} \Lambda^k R_y.
 \end{aligned}$$

By the strict inequalities between the eigenvalues we have

$$(\Lambda^k L_y \Lambda^{-k})_{ij} = \begin{cases} 1 & i = j \\ \left(\frac{\lambda_i}{\lambda_j}\right)^k \ell_{ij} & i > j \\ 0 & i < j. \end{cases}$$

Hence $\Lambda^k L_y \Lambda^{-k} \rightarrow I$ (because $\frac{|\lambda_i|}{|\lambda_j|} < 1$ if $i > j$). Hence with

$$A_1^k = L_x R_x (\Lambda^k L_y \Lambda^{-k}) \Lambda^k R_y$$

and

$$A_1^k = P_k Q_k$$

we conclude that $\lim_{k \rightarrow \infty} P_k = L_x$. Therefore

$$L_k = P_{k-1}^{-1} P_k \rightarrow I.$$

Finally we have that A_k must be upper triangular because

$$L_k^{-1} A_k = R_k$$

is upper triangular. □

This exposes all the eigenvalues of A . Therefore the eigenvectors can be determined.

7.3 The QR algorithm

Certain numerical problems with the LU algorithm have led to the QR algorithm, which is based on the decomposition of the matrix A as

$$A = QR$$

where Q is unitary and R is upper triangular.

Theorem 7.3.1 (QR-factorization). *(i) Suppose A is in $M_{n,m}$ and $n \geq m$. Then there is a matrix $Q \in M_{n,m}$ with orthogonal columns and an upper triangular matrix $R \in M_m$ such that $A = QR$.*

(ii) If $n = m$, then Q is unitary. If A is nonsingular the diagonal entries of R can be chosen to be positive.

(iii) If A is real; then Q and R may be chosen to be real.

Proof. (i) We proceed inductively. Let $a_1, \dots, a_n$ denote the columns of A and $q_1, q_2, \dots, q_m$ denote the columns of Q . The basic idea of the QR-factorization is to orthogonalize the columns of A from left to right. Then the columns can be expressed by the formulas $a_k = \sum_{i=1}^k c_{ik} q_i$, $k = 1, \dots, n$. The coefficients of the expansion become, respectively, the entries of the k^{th} column of R , completed by $n - k$ zeros. (Of course, if the rank of A is less than m , we fill in arbitrary orthogonal vectors which we know exist as $m \leq n$.) For the details, first define $q_1 = a_1 / \|a_1\|$. To compute q_2 we use the Gram-Schmidt procedure.

$$\begin{aligned}\hat{q}_2 &= a_2 - \langle q_1, a_2 \rangle q_1 \\ q_2 &= \hat{q}_2 / \|\hat{q}_2\|.\end{aligned}$$

Tracing backwards note that

$$\begin{aligned}a_2 &= \hat{q}_2 + \langle q_1, a_2 \rangle q_1 \\ &= \|q_2\| q_2 + \langle q_1, a_2 \rangle q_1.\end{aligned}$$

So we have

$$\begin{bmatrix} a_1 & a_2 & a_3 & \dots \\ \downarrow & \downarrow & \downarrow & \dots \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 & \dots \\ \downarrow & \downarrow & \downarrow & \dots \end{bmatrix} \begin{bmatrix} \|a_1\| & \langle q_1, a_2 \rangle & \dots \\ 0 & \|q_2\| & \dots \\ \vdots & 0 & \dots \\ 0 & 0 & \dots \end{bmatrix}.$$

Instead of the full inductive step we compute q_3 and finish at that point

$$\begin{aligned}\hat{q}_3 &= a_3 - \langle q_1, a_3 \rangle q_1 - \langle q_2, a_3 \rangle q_2 \\ q_3 &= \hat{q}_3 / \|\hat{q}_3\|.\end{aligned}$$

Hence

$$a_3 = \|q_3\| q_3 + \langle q_1, a_3 \rangle q_1 + \langle q_2, a_3 \rangle q_2.$$

The third column of R is thus given by

$$r_3 = [\langle q_1, a_3 \rangle, \langle q_2, a_3 \rangle, \|\hat{q}_3\|, 0, 0, \dots, 0]^T.$$

In this way we see that the columns of Q are orthogonal and the matrix R is upper triangular, with an exception. That is the possibility that $\hat{q}_k = 0$ for some k . In this degenerate case we take q_k to be any vector orthogonal to the span of $a_1, a_2, \dots, a_m$, and we take $r_{kj} = 0$, $j = k, k+1, \dots, m$. Also we note that if $\hat{q}_k = 0$, then a_k is linearly dependent on $a_1, a_2, \dots, a_{k-1}$, and hence on $q_1, q_2, \dots, q_{k-1}$. Select the coefficients $r_{1k}, \dots, r_{k-1k}$ to reflect this dependence.

- (ii) If $m = n$, the process above yields a unitary matrix. If A is nonsingular, the process above yields a matrix R with a positive diagonal.
- (iii) If A is a real, all operators above can be carried out in real arithmetic.

□

Now what about the uniqueness of the decomposition? Essentially the uniqueness is true up to a multiplication by a diagonal matrix, except in the case when the matrix has rank is less than m , when there is no form of uniqueness. Suppose that the rank of A is m .

Then application of the Gram-Schmidt procedure yields a matrix R with positive diagonal. Suppose that A has two QR factorizations, QR and PS with upper triangular factors having positive diagonals. Then

$$P^*Q = SR^{-1}$$

We have that SR^{-1} is upper triangular and moreover has a positive diagonal. Also, P^*Q is unitary. We know that the only upper triangular unitary matrices are diagonal matrices, and finally the only unitary matrix with a positive diagonal is the identity matrix. Therefore $P^*Q = I$, which is to say that $P = Q$. We summarize as

Corollary 7.3.1. *Suppose A is in $M_{n,m}$ and $n \geq m$. If $\text{rank}(A) = m$ then the QR factorization of $A = QR$ with upper triangular matrix R having a positive diagonal is unique.*

The QR algorithm

The QR algorithm parallels the LR algorithm almost identically. Suppose A is in M_n . Define

$$\begin{aligned} A_1 &= Q_1 R_1 \\ A_2 &\equiv R_1 Q_1. \end{aligned}$$

Also

$$Q_1^* A_1 Q_1 = A_2.$$

Then decompose A_2 into a QR decomposition

$$A_2 = Q_2 R_2$$

and

$$Q_2^* A_2 Q_2 = R_2 Q_2 \equiv A_3.$$

Also

$$Q_2^* Q_1^* A_1 Q_1 Q_2 = R_2 Q_2 = A_3.$$

Proceed sequentially

$$\begin{aligned} A_k &= Q_k R_k \\ A_{k+1} &= R_k Q_k \\ Q_k^* A_k Q_k &= A_{k+1}. \end{aligned}$$

Let

$$\begin{aligned} P_k &= Q_1 Q_2 \dots Q_k \\ T_k &= R_k R_{k-1} \dots R_1. \end{aligned}$$

Then

$$P_k^* A_1 P_k = A_{k+1}.$$

whence

$$P_k A_{k+1} = A_1 P_k.$$

Also we have

$$\begin{aligned}
 P_k T_k &= P_{k-1} Q_k R_k T_{k-1} \\
 &= P_{k-1} A_k T_{k-1} \\
 &= A_1 P_{k-1} T_{k-1} \\
 &= \dots \\
 &= A_1^k.
 \end{aligned}$$

Theorem 7.3.2. *Let $A \in M_n$ be given, and assume the eigenvalues of A satisfy*

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0.$$

Then the iterations A_k converge to a triangular matrix.

Proof. Our hypothesis gives that A is diagonalizable, and we write $A \sim \Lambda = \text{diag}(\lambda_1 \dots \lambda_n)$. That is,

$$A_1 = S \Lambda S^{-1}$$

where $\Lambda = \text{diag}(\lambda_1 \dots \lambda_n)$. Let

$$\begin{array}{l|l}
 X = S = Q_x R_x & \text{here } QR \\
 Y = S^{-1} = L_y U_y & \text{here } LU.
 \end{array}$$

Then

$$\begin{aligned}
 A_1^k &= Q_x R_x \Lambda^k L_y U_y \\
 &= Q_x R_x \Lambda^k L_y \Lambda^{-k} \Lambda^k U_y \\
 &= Q_x (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_y
 \end{aligned}$$

where

$$\begin{aligned}
 E_k &= \Lambda^k L_y \Lambda^{-k} - I \\
 (E_k)_{ij} &= \begin{cases} 0 & i = j \\ (\lambda_i / \lambda_j)^k \ell_{ij} & i > j \\ 0 & i < j. \end{cases}
 \end{aligned}$$

It follows that $I + R_x E_k R_x^{-1} \rightarrow I$, and $R_x \Lambda^{-k} U_y$ is upper triangular. Thus

$$Q_x (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_y = P_k T_k.$$

The matrix $I + R_x E_k R_x^{-1}$ can be QR factored as $\tilde{U}_k \tilde{R}_k$, and since $I + R_x E_k R_x^{-1} \rightarrow I$, it follows that we can assume both $\tilde{U}_k \rightarrow I$ and $\tilde{R}_k \rightarrow I$. Hence

$$A_1^k = Q_x \tilde{U}_k [\tilde{R}_k (I + R_x E_k R_x^{-1}) R_x \Lambda^k U_y] = P_k T_k.$$

with the first factor unitary and the second factor upper triangular. Since we have assumed (by the eigenvalue condition) that A is nonsingular, this factorization is essentially unique, where possibly a multiplication by a diagonal matrix must be applied to give the upper triangular factor on the right a positive diagonal. Just what is the form of the diagonal matrix can be seen from the following. Let $\Lambda = |\Lambda| \Lambda_1$, where $|\Lambda|$ is the diagonal matrix of moduli of the elements of Λ and where Λ_1 is the unitary matrix of the signs of each eigenvalue respectively. We also take $U_y = \Lambda_2 (\Lambda_2^* U_y)$ where Λ_2 is a unitary matrix chosen so that $\Lambda_2^* U$ has a positive diagonal. Then

$$A_1^k = Q_x \tilde{U}_k \Lambda_2 \Lambda_1^k [(\Lambda_2 \Lambda_1^k)^{-1} \tilde{R}_k (I + R_x E_k R_x^{-1}) R_x (\Lambda_2 \Lambda_1^k) |\Lambda|^k (\Lambda_2^* U_y)] = P_k T_k.$$

From this we obtain P_k is essentially asymptotic to $Q_x \tilde{U}_k \Lambda_2 \Lambda_1^k$ and from this we obtain that

$$Q_k = P_{k-1}^{-1} P_k \rightarrow \Lambda_1$$

which is diagonal. Finally, it follows that A_k is upper triangular since

$$Q_k^{-1} A_k = R_k$$

In the limit therefore A is similar to an upper triangular matrix. $\square$

Example 7.3.1. Apply the QR method to the matrix

$$A := \begin{bmatrix} 2.3 & 1 & 2 \\ 2 & 2 & 2.1 \\ 3 & 2 & 0 \end{bmatrix}$$

The matrix A has eigenvalues 5.45, 0.723, -1.87 . The successive iterations are

$$\begin{aligned}
A_2 &= \begin{bmatrix} 5.10 & -0.511 & 2.13 \\ 0.631 & 0.662 & 0.136 \\ 1.42 & -0.0202 & -1.44 \end{bmatrix} & A_3 &= \begin{bmatrix} 5.51 & -1.02 & -0.36 \\ -0.0146 & 0.666 & 0.482 \\ 0.513 & 0.240 & -1.84 \end{bmatrix} \\
A_4 &= \begin{bmatrix} 5.46 & -1.41 & 0.482 \\ -0.0372 & 0.495 & 0.672 \\ 0.169 & 0.815 & -1.62 \end{bmatrix} & A_5 &= \begin{bmatrix} 5.47 & -0.366 & -1.26 \\ -0.0404 & -0.462 & 1.39 \\ 0.0430 & 1.21 & -0.677 \end{bmatrix} \\
A_6 &= \begin{bmatrix} 5.46 & -1.13 & -0.687 \\ -0.0184 & -1.52 & 0.813 \\ 0.00826 & 0.983 & 0.381 \end{bmatrix} & A_7 &= \begin{bmatrix} 5.45 & 0.529 & -1.18 \\ -0.00682 & -1.78 & 0.585 \\ 0.00115 & 0.414 & 0.638 \end{bmatrix} \\
A_8 &= \begin{bmatrix} 5.43 & 0.684 & -1.09 \\ -0.000822 & -1.87 & 0.229 \\ 0.0000215 & 0.0659 & 0.729 \end{bmatrix}
\end{aligned}$$

Note the gradual appearance of the eigenvalues on the diagonal.

Remark. These iterations were carried out in precision 3 arithmetic, which affects the rate of convergence to triangular form.

7.4 Least Squares

As we know, if $A \in M_{n,m}$ with $m < n$ it is generally not possible to solve the overdetermined system

$$Ax = b.$$

For example, suppose we have the data $\{(x_i, y_i)\}_{i=1}^n$, with the x -coordinates distinct. We may wish to “fit” a straight to this data. This means we want to find coefficients m and b so that

$$b + mx_i = y_i, \quad i = 1, \dots, n. \quad (\star)$$

Taking the matrix and data vector

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and $z = [b, m]^T$, the system $(\star)$ becomes $Az = b$. Usually $n \gg 2$. Hence there is virtually no hope to determine a unique solution to system.

However, there are numerous ways to determine constants m and b so that the resulting line represents the data. For example, owing to the distinctness of the x -coordinates, it is possible to solve any 2×2 subsystem of

$Az = b$. Other variations exist. A new 2×2 system could be created by creating two averages of the data, say left and right, and solving. Assume the sequence $\{x_j\}$ is ordered from least to greatest. Define $x_\ell = \frac{1}{k} \sum_{j=1}^k x_j$ and $x_r = \frac{1}{n-k} \sum_{j=k+1}^n x_j$. Let y_ℓ and y_r denote the corresponding averages for the ordinates. Then define the intercept b and slope m by solving the system

$$\begin{bmatrix} 1 & x_\ell \\ 1 & x_r \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} y_\ell \\ y_r \end{bmatrix}$$

While this will normally give a reasonable approximating line, its value has little utility beyond its naive simplicity and visual appearance. What is desired is to establish a criteria for choosing the line.

Define the **residual** of the approximation $r = b - Az$. It makes perfect sense to consider finding $z = [b, m]^T$ for which the residual is minimized in some norm. Any norm can be selected here, but on practical grounds the best norm to use is the Euclidean norm $\|\cdot\|_2$. The vector Az that yields the minimal norm residual is the one for which $(b - Az) \perp Aw$, for we are seeking the nearest value in the Aw to the vector b . It can be found by select the one for the solution, Az , for which

$$b - Ax \perp Aw \quad \text{all } w.$$

This means

$$\langle b - Ax, Ay \rangle = 0 \quad \text{all } y$$

or

$$\langle A^T(b - Ay), y \rangle = 0 \quad \text{all } y$$

or

$$A^T(b - Ay) = 0$$

$A^T Ay = A^T b.$	Normal Equations
-------------------	---------------------

The **least squares** solution to $Ax = b$ is given by the solution to the normal equation

$$A^T Ay = A^T b.$$

Suppose we have the QR decomposition for A . Then if A is real

$$\begin{aligned} A^T A &= R^T Q^T Q R = R^T R \\ A^T y &= R^T Q y. \end{aligned}$$

Hence the normal equations become

$$R^T R x = R^T Q y.$$

Assuming that the rank of A is m , we must have that R and hence R^T is invertible. Therefore we have the least squares solution is given by the triangular system

$$R x = Q y.$$

7.5 Exercises

1. If $A \in M(\mathbb{C})$ has rank k , show that there is a permutation matrix P such that PA has its first k principal determinants nonzero.
2. For the least squares fit of a straight line determine R and Q .
3. In the case of data

$$A^T A = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \quad A^T b = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}.$$

4. In attempting to solve a quadratic fit we have the model

$$c + bx_i + ax_i^2 = y_i \quad i = 1, \dots, n.$$

The system is

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

The normal equations have the matrix and data given by

$$A^T A = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \quad A^T b = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}.$$

5. Find the normal equations for the least squares fit of data to a polynomial of degree k .

Chapter 8

Jordan Normal Form

8.1 Minimal Polynomials

Recall $p_A(x) = \det(xI - A)$ is called the characteristic polynomial of the matrix A .

Theorem 8.1.1. *Let $A \in M_n$. Then there exists a unique monic polynomial $q_A(x)$ of minimum degree for which $q_A(A) = 0$. If $p(x)$ is any polynomial such that $p(A) = 0$, then $q_A(x)$ divides $p(x)$.*

Proof. Since there is a polynomial $p_A(x)$ for which $p_A(A) = 0$, there is one of minimal degree, which we can assume is monic. by the Euclidean algorithm

$$p_A(x) = q_A(x)h(x) + r(x)$$

where $\deg r(x) < \deg q_A(x)$. We know

$$p_A(A) = q_A(A)h(A) + r(A).$$

Hence $r(A) = 0$, and by the minimality assumption $r(x) \equiv 0$. Thus q_A divided $p_A(x)$ and also any polynomial for which $p(A) = 0$. to establish that q_A is unique, suppose $q(x)$ is another monic polynomial of the same degree for which $q(A) = 0$. Then

$$r(x) = q(x) - q_A(x)$$

is a polynomial of degree less than $q_A(x)$ for which $r(A) = q(A) - q_A(A) = 0$. This cannot be unless $r(x) \equiv 0 = 0q = q_A$. $\square$

Definition 8.1.1. The polynomial $q_A(x)$ in the theorem above is called the minimal polynomial.

Corollary 8.1.1. *If $A, B \in M_n$ are similar, then they have the same minimal polynomial.*

Proof.

$$\begin{aligned} B &= S^{-1}AS \\ q_A(B) &= q_A(S^{-1}AS) = S^{-1}q_A(A)S = q_A(A) = 0. \end{aligned}$$

If there is a minimal polynomial for B of smaller degree, say $q_B(x)$, then $q_B(A) = 0$ by the same argument. This contradicts the minimality of $q_A(x)$. $\square$

Now that we have a minimum polynomial for any matrix, can we find a matrix with a given polynomial as its minimum polynomial? Can the degree of the polynomial and the size of the matrix match? The answers to both questions are affirmative and presented below in one theorem.

Theorem 8.1.2. *For any n^{th} degree polynomial*

$$p(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0$$

there is a matrix $A \in M_n(\mathbb{C})$ for which it is the minimal polynomial.

Proof. Consider the matrix given by

$$A = \begin{bmatrix} 0 & 0 & \cdots & \cdots & -a_0 \\ 1 & 0 & & & -a_1 \\ 0 & 1 & 0 & \ddots & -a_2 \\ & \ddots & & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{bmatrix}.$$

Observe that

$$\begin{aligned} Ae_1 &= e_1 = A^0e_1 \\ Ae_1 &= e_2 = Ae_1 \\ Ae_2 &= e_3 = A^2e_1 \\ &\vdots \\ Ae_{n-1} &= e_n = A^{n-1}e_1 \end{aligned}$$

and

$$Ae_n = -a_{n-1}e_n - a_{n-2}e_{n-1} - \cdots - a_1e_2 - a_0e_1$$

Since $Ae_n = A^n e_1$, it follows that

$$p(A)e_1 = A^n e_1 + a_{n-1}A^{n-1}e_1 + a_{n-2}A^{n-2}e_1 + \cdots + a_1Ae_1 + a_0Ie_1 = 0$$

Also

$$p(A)e_k = p(A)A^{k-1}e_1 = A^{k-1}p(A)e_1 = A^{k-1}(0) = 0 \quad k = 2, \dots, n.$$

Hence $p(A)e_j = 0$ for $j = 1 \dots n$. Thus $p(A) = 0$. We know also that $p(x)$ is monic. Suppose now that

$$q(x) = x^m + b_{m-1}x^{m-1} \cdots + b_1x + b_0$$

where $m < n$ and $q(A) = 0$. Then

$$\begin{aligned} q(A)e_1 &= A^m e_1 + b_{m-1}A^{m-1}e_1 + \cdots + b_1Ae_1 + b_0e_1 \\ &= e_{m+1} + b_{m-1}e_m + \cdots + b_1e_2 + b_0e_1 = 0. \end{aligned}$$

But the vectors $e_{m+1} \dots e_1$ are linear independent from which we conclude that $q(A) = 0$ is impossible. Thus $p(x)$ is minimal. $\square$

Definition 8.1.2. For a given monic polynomial $p(x)$, the matrix A constructed above is called the *companion* matrix to p .

The transpose of the companion matrix can also be used to generate a linear differential system which has the same characteristic polynomial as a given n^{th} order differential equation. Consider the linear differential equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1y' + a_0 = 0.$$

This n^{th} order ODE can be converted to a first order system as follows:

$$\begin{aligned} u_1 &= y \\ u_2 &= u'_1 &= y' \\ u_3 &= u'_2 &= y'' \\ &\vdots &\vdots \\ u_n &= u'_{n-1} &= y^{(n-1)} \end{aligned}$$

Then we have

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}' = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & 0 \\ & & 0 & 1 & \\ & & & \ddots & \ddots \\ & 0 & & & \ddots & 1 \\ a_0 & -a_1 & \dots & \dots & -a_{n-1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

8.2 Invariant subspaces

There seems to be no truly simple way to the Jordan normal form. The approach taken here is intended to reveal a number of features of a matrix, interesting in their own right. In particular, we will construct “generalized eigenspaces” that envelop the entire connection of a matrix with its eigenvalues. We have in various ways considered subspaces V of $\mathbb{C}_n$ that are invariant under the matrix $A \in M_n(\mathbb{C})$. Recall this means that $AV \subset V$. For example, eigenvectors can be used to create invariant subspaces. Null spaces, the eigenspace of the zero eigenvalue, are invariant as well. Triangular matrices furnish an easily recognizable sequence of invariant subspaces. Assuming $T \in M_n(\mathbb{C})$ is upper triangular, it is easy to see that the subspaces generated by the coordinate vectors $\{e_1, \dots, e_m\}$ for $m = 1, \dots, n$ are invariant under T .

We now consider a specific type of invariant subspace that will lead the so-called Jordan normal form of a matrix, the closest matrix similar to A that resembles a diagonal matrix.

Definition 8.2.1 (Generalized Eigenspace). Let $A \in M_n(C)$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$. Define the *generalized eigenspace* pertaining to λ_i by

$$V_{\lambda_i} = \{x \in \mathbb{C}_n \mid (A - \lambda_i I)^n x = 0\}$$

Observe that all the eigenvectors pertaining to λ_i are contained in V_{λ_i} . If the span of the eigenvectors pertaining to λ_i is not equal to V_{λ_i} then there must be a positive power p and a vector x such that $(A - \lambda_i I)^p x = 0$ but that $y = (A - \lambda_i I)^{p-1} x \neq 0$. Thus y is an eigenvector pertaining to λ_i . For this reason we will call V_{λ_i} the space of **generalized eigenvectors** pertaining to λ_i . Our first result, that V_{λ_i} is invariant under A , is simple to prove, noting that only closure under vector addition and scalar multiplication need be established.

Theorem 8.2.1. *Let $A \in M_n(C)$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$. Then for each $i = 1, \dots, k$, V_{λ_i} is an invariant subspace of A .*

One might question as to whether V_{λ_i} could be enlarged by allowing higher powers than n in the definition. The negative answer is most simply expressed by evoking the Hamilton-Cayley theorem. We write the characteristic polynomial $p_A(\lambda) = \prod (\lambda - \lambda_i)^{m_A(\lambda_i)}$, where $m_A(\lambda_i)$ is the algebraic multiplicity of λ_i . Since $p_A(A) = \prod (A - \lambda_i I)^{m_A(\lambda_i)} = 0$, it is an easy matter to see that we exhaust all of $\mathbb{C}_n$ with the spaces V_{λ_i} . This is to say that allowing higher powers in the definition will not increase the subspaces V_{λ_i} . Indeed, as we shall see, the power of $(\lambda - \lambda_i)$ can be decreased to the geometric multiplicity $m_g(\lambda_i)$ the power for λ_i . For now the general power n will suffice. One very important result, and an essential first step in deriving the Jordan form, is to establish that any square matrix A is similar to a block diagonal matrix, with each block carrying a single eigenvalue.

Theorem 8.2.2. *Let $A \in M_n(C)$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ and with invariant subspaces V_{λ_i} , $i = 1, 2, \dots, k$. Then (i) The spaces V_{λ_i} , $j = 1, \dots, k$ are mutually linearly independent. (ii) $\bigoplus_{i=1}^k V_{\lambda_i} = \mathbb{C}_n$ (alternatively $\mathbb{C}_n = \mathfrak{S}(V_{\lambda_1}, \dots, V_{\lambda_k})$) (iii) $\dim V_{\lambda_i} = m_A(\lambda_i)$. (iv) A is similar to a block diagonal matrix with k blocks $A_1, \dots, A_k$. Moreover, $\sigma(A_i) = \{\lambda_i\}$ and $\dim A_i = m_A(\lambda_i)$.*

Proof. (i) It should be clear that the subspaces V_{λ_i} are linearly independent of each other. For if there is a vector x in both V_{λ_i} and V_{λ_j} then there is a vector for some integer q , it must be true that $(A - \lambda_j I)^{q-1} x \neq 0$ but $(A - \lambda_j I)^q x = 0$. This means that $y = (A - \lambda_j I)^{q-1} x$ is an eigenvector pertaining to λ_j . Since $(A - \lambda_i I)^n x = 0$ we must also have that

$$\begin{aligned} (A - \lambda_j I)^{q-1} (A - \lambda_i I)^n x &= (A - \lambda_i I)^n (A - \lambda_j I)^{q-1} x \\ &= (A - \lambda_i I)^n y = 0 \\ &= \sum_{k=0}^n \binom{n}{k} (-\lambda_i)^{n-k} A^k y \\ &= \sum_{k=0}^n \binom{n}{k} (-\lambda_i)^{n-k} \lambda_j^k y \\ &= (\lambda_j - \lambda_i)^n y = 0 \end{aligned}$$

This is impossible unless $\lambda_j = \lambda_i$. (ii) The key part of the proof is to *block* diagonalize A with respect to these invariant subspaces. To that end, let S

be the matrix with columns generated from bases of the individual V_{λ_i} taken in the order of the indices. Supposing there are more linearly independent vectors in $\mathbb{C}_n$ other than those already selected, fill out the matrix S with vectors linearly independent to the subspaces V_{λ_i} , $i = 1, \dots, k$. Now define $\tilde{A} = S^{-1}AS$. We conclude by the invariance of the subspaces and their mutual linear independence that $\tilde{A}$ has the following block structure.

$$\tilde{A} = S^{-1}AS = \begin{bmatrix} A_1 & 0 & \cdots & 0 & * \\ 0 & A_2 & & 0 & * \\ \vdots & & \ddots & & \vdots \\ & & & A_k & * \\ 0 & \cdots & & 0 & B \end{bmatrix}$$

It follows that

$$p_A(\lambda) = p_{\tilde{A}}(\lambda) = \left(\prod p_{A_i}(\lambda) \right) p_B(\lambda)$$

Any root r of $p_B(\lambda)$ must be an eigenvalue of A , say λ_j , and there must be an eigenvector x pertaining to λ_j . Moreover, due to the block structure we can assume that $x = [0, \dots, 0, x]^T$, where there are k blocked zeros of the sizes of the A_i respectively. Then it is easy to see that $ASx = \lambda_j Sx$, and this implies that $Sx \in V_{\lambda_j}$. Thus there is another vector in V_{λ_j} , which contradicts its definition. Therefore B is null, or what is the same thing, $\oplus_{i=1}^k V_{\lambda_i} = \mathbb{C}_n$. (iii) Let $d_i = \dim V_{\lambda_i}$. From (ii) know $\sum d_i = n$. Suppose that $\lambda_i \in \sigma(A_j)$. Then there is another eigenvector x pertaining to λ_i and for which $A_j x = \lambda_i x$. Moreover, this vector has the form $x = [0, \dots, 0, x, 0, \dots, 0]^T$, analogous to the argument above. By construction $Sx \notin V_{\lambda_i}$, but $ASx = \lambda_i Sx$, and this contradicts the definition of V_{λ_i} . We thus have that $p_{A_i}(\lambda) = (\lambda - \lambda_i)^{d_i}$. Since $p_A(\lambda) = \prod p_{A_i}(\lambda) = \prod (\lambda - \lambda_i)^{m_A(\lambda_i)}$, it follows that $d_i = m_A(\lambda_i)$. (iv) Putting (ii), and (iii) together gives the block diagonal structure as required. $\square$

On account of the mutual linear independence of the invariant subspaces V_{λ_i} and the fact that they exhaust $\mathbb{C}_n$ the following corollary is immediate.

Corollary 8.2.1. *Let $A \in M_n(\mathbb{C})$ with spectrum $\sigma(A) = \lambda_1, \dots, \lambda_k$ and with generalized eigenspaces V_{λ_i} , $i = 1, 2, \dots, k$. Then each $x \in \mathbb{C}_n$ has a unique representation $x = \sum_{i=1}^k x_i$ where $x_i \in V_{\lambda_i}$.*

Another interesting result which reveals how the matrix works as a linear transformation is to decompose the it into components with respect to the

generalized eigenspaces. In particular, viewing the block diagonal form

$$\tilde{A} = S^{-1}AS = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & 0 \\ \vdots & & \ddots & \\ & & & A_k \end{bmatrix}$$

the space $\mathbb{C}_n$ can be split into a direct sum of subspaces $E_1, \dots, E_k$ based on coordinate blocks. This is accomplished in such that any vector $y \in \mathbb{C}_n$ can be written uniquely as $y = \sum_{i=1}^k y_i$ where the $y_i \in E_i$. (Keep in mind that each $y_i \in \mathbb{C}_n$; its coordinates are zero outside the coordinate block pertaining E_i .) Then $\tilde{A}y = \tilde{A} \sum_{i=1}^k y_i = \sum_{i=1}^k \tilde{A}y_i = \sum_{i=1}^k A_i y_i$. This provides a computational tool — when this block diagonal form is known. Note that the blocks correspond directly to the invariant subspaces by $SE_i = V_{\lambda_i}$. We can use these invariant subspaces to get at the minimal polynomial. For each $i = 1, \dots, k$ define

$$m_i = \min_j \{(A - \lambda_i I)^j x = 0 \mid x \in V_{\lambda_i}\}$$

Theorem 8.2.3. *Let $A \in M_n(\mathbb{C})$ with spectrum $\sigma(A) = \lambda_1, \dots, \lambda_k$ and with invariant subspaces V_{λ_i} , $i = 1, 2, \dots, k$. Then the minimal polynomial of A is given by*

$$q(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{m_i}$$

Proof. Certainly we see that for any vector $x \in V_{\lambda_j}$

$$q(A)x = \left(\prod_{i=1}^k (A - \lambda_i I)^{m_i} \right) x = 0$$

Hence, the minimal polynomial $q_A(\lambda)$ divides $q(A)$. To see that indeed they are in fact equal, suppose that the minimal polynomial has the form

$$q_A(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{\hat{m}_i}$$

where $\hat{m}_i \leq m_i$, for $i = 1, \dots, k$ and in particular $\hat{m}_j < m_j$. By construction there must exist a vector $x \in V_{\lambda_j}$ such that $(A - \lambda_j I)^{m_j} x = 0$ but $y =$

$(A - \lambda_j I)^{m_j-1} x \neq 0$. Then if

$$\begin{aligned} q(A)x &= \left(\prod_{i=1}^k (A - \lambda_i I)^{\hat{m}_i} \right) x \\ &= \left(\prod_{\substack{i=1 \\ i \neq j}}^k (A - \lambda_i I)^{\hat{m}_i} \right) y \\ &= 0 \end{aligned}$$

This cannot be because the contrary implies that there is another vector in one of the invariant subspaces V_{λ_k} . $\square$

Just one more step is needed before the Jordan normal form can be derived. For a given V_{λ_i} we can interpret the spaces in a heirarchical viewpoint. We know that V_{λ_i} contains all the eigenvectors pertaining to λ_i . Call these eigenvectors the *first order generalized eigenvectors*. If the span of these is not equal to V_{λ_i} , then there must be a vector $x \in V_{\lambda_i}$ for which $y = (A - \lambda_i I)^2 x = 0$ but $(A - \lambda_i I)x \neq 0$. That is to say y is an eigenvector of A pertaining to λ_i . Call such vectors *second order generalized eigenvectors*. In general we call an $x \in V_{\lambda_i}$ a generalized eigenvector of order j if $y = (A - \lambda_i I)^j x = 0$ but $(A - \lambda_i I)^{j-1} x \neq 0$. In light of our previous discussion V_{λ_i} contains generalized eigenvectors of order up to but not greater than m_{λ_i} .

Theorem 8.2.4. *Let $A \in M_n(\mathbb{C})$ with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_k\}$ and with invariant subspaces V_{λ_i} , $i = 1, 2, \dots, k$.*

(i) *Let $x \in V_{\lambda_i}$ be a generalized eigenvector of order p . Then the vectors*

$$x, (A - \lambda_i I)x, (A - \lambda_i I)^2 x, \dots, (A - \lambda_i I)^{p-1} x \quad (1)$$

are linearly independent.

(ii) *The subspace of $\mathbb{C}_n$ generated by the vectors in (1) is an invariant subspace of A .*

Proof. (i) To prove linear independence of a set of vectors we suppose linear dependence. That is there is a smallest integer k and constants b_j such that

$$\sum_{j=0}^k b_j (A - \lambda_i I)^j x = 0$$

where $b_k \neq 0$. Solving we obtain $b_k (A - \lambda_i I)^k x = -\sum_{j=0}^{k-1} b_j (A - \lambda_i I)^j x$. Now apply $(A - \lambda_i I)^{p-k}$ to both sides and obtain a new linearly dependent set as the following calculation shows.

$$\begin{aligned} 0 &= b_k (A - \lambda_i I)^{p-k+k} x = -\sum_{j=0}^{k-1} b_j (A - \lambda_i I)^{j+p-k} x \\ &= -\sum_{j=p-k}^{p-1} b_{j+p-k} (A - \lambda_i I)^j x \end{aligned}$$

The key point to note here is that the lower limit of the sum is increased. This new linearly dependent set, which we denote with the notation $\sum_{j=p-k}^{p-1} c_j (A - \lambda_i I)^j x$ can be split in the same way as before, where we assume with no loss in generality that $c_{p-1} \neq 0$. Then

$$c_{p-1} (A - \lambda_i I)^{p-1} x = -\sum_{j=p-k}^{p-2} c_j (A - \lambda_i I)^j x$$

Apply $(A - \lambda_i I)$ to both sides to get

$$\begin{aligned} 0 &= c_{p-1} (A - \lambda_i I)^p x = -\sum_{j=p-k}^{p-2} c_j (A - \lambda_i I)^{j+1} x \\ &= -\sum_{j=p-k+1}^{p-1} c_{j-1} (A - \lambda_i I)^j x \end{aligned}$$

Thus we have obtained another linearly independent set with the lower limit of powers increased by one. Continue this process until the linear dependence of $(A - \lambda_i I)^{p-1} x$ and $(A - \lambda_i I)^{p-2} x$ is achieved. Thus we have

$$\begin{aligned} c (A - \lambda_i I)^{p-1} x &= d (A - \lambda_i I)^{p-2} x \\ (A - \lambda_i I) y &= \frac{d}{c} y \end{aligned}$$

where $y = (A - \lambda_i I)^{p-2} x$. This implies that $\lambda_i + \frac{d}{c}$ is a new eigenvalue with eigenvector $y \in V_{\lambda_i}$, and of course this is a contradiction. (ii) The invariance under A is more straightforward. First note that while $x_1 = x$,

$x_2 = (A - \lambda I)x = Ax - \lambda x$ so that $Ax = x_2 + \lambda x_1$. Consider any vector y defined by $y = \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j x$. It follows that

$$\begin{aligned}
 Ay &= A \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j x \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j Ax \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j (x_2 + \lambda x_1) \\
 &= \sum_{j=0}^{p-1} b_j (A - \lambda_i I)^j [(A - \lambda_i I)x_1 + \lambda x_1] \\
 &= \sum_{j=0}^{p-1} c_j (A - \lambda_i I)^j x
 \end{aligned}$$

where $c_j = b_{j-1} - \lambda$ for $j > 0$ and $c_0 = -\lambda$, which proves the result. $\square$

8.3 The Jordan Normal Form

We need a lemma that points in the direction we are headed, that being the use of invariant subspaces as a basis for the (Jordan) block diagonalization of any matrix. These results were discussed in detail in the Section 8.2. The restatement here illustrates the “invariant subspace” nature of the result, irrespective of generalized eigenspaces. Its proof is elementary and is left to the reader.

Lemma 8.3.1. *Let $A \in M_n(\mathbb{C})$ with invariant subspace $V \subset \mathbb{C}_n$.*

(i) Suppose $v_1 \dots v_k$ is a basis for V and S is an invertible matrix with the first k columns given by $v_1 \dots v_k$. Then

$$S^{-1}AS = \left[\begin{array}{c|c} & \\ \hline * & * \\ 0 & * \end{array} \right]^{1 \dots k}.$$

(ii) Suppose that $V_1, V_2 \subset \mathbb{C}_n$ are two invariant subspaces of A and $\mathbb{C}_n = V_1 \oplus V_2$. Let the (invertible) matrix S consist respectively of bases from V_1 and V_2 as its columns. Then

$$S^{-1}AS = \left[\begin{array}{c|c} * & 0 \\ \hline 0 & * \end{array} \right].$$

Definition 8.3.1. Let $\lambda \in \mathbb{C}$. A Jordan block $J_k(\lambda)$ is a $k \times k$ upper triangular matrix of the form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ 0 & & \ddots & 1 \\ & & & \lambda \end{bmatrix}.$$

A Jordan matrix is any matrix of the form

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_k}(\lambda_k) \end{bmatrix}.$$

where the matrices J_{n_i} are Jordan blocks. If $J \in M_n(\mathbb{C})$, then $n_1 + n_2 + \cdots + n_k = n$.

Theorem 8.3.1 (Jordan normal form). Let $A \in M_n(\mathbb{C})$. Then there is a nonsingular matrix $S \in M_n$ such that

$$A = S \begin{bmatrix} J_{n_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{n_k}(\lambda_k) \end{bmatrix} S^{-1} = SJS^{-1}$$

where $J_{n_i}(\lambda_i)$ is a Jordan block, where $n_1 + n_2 + \cdots + n_k = n$. J is unique up to permutations of the blocks. The eigenvalues $\lambda_1, \dots, \lambda_k$ are not necessarily distinct. If A is real with real eigenvalues, then S can be taken as real.

Proof. This result is proved in four steps.

- (1) Block diagonalize (by similarity) into invariant subspaces pertaining to $\sigma(A)$. This is accomplished as follows. First block diagonalize the matrix according to the generalized eigenspaces $V_{\lambda_i} = \{x \in \mathbb{C}_n \mid (A - \lambda_i I)^n x =$

$0\}$ as discussed in the previous section. Beginning with the highest order eigenvector in $x \in V_{\lambda_i}$, construct the invariant subspace as in (1) of Section 8.2. Repeat this process until all generalized eigenvectors have been included in an invariant subspace. This includes of course first order eigenvectors that are not associated with higher order eigenvectors. These invariant subspaces have dimension one. Each of these invariant subspaces is linearly independent from the others. Continue this process for all the generalized eigenspaces. This exhausts $\mathbb{C}^n$. Each of the blocks contains exactly one eigenvector. The dimensions of these invariant subspaces can range from one to m_{λ_i} , there being at least one subspace of dimension m_{λ_i} .

- (2) Triangularize each block by Schur's theorem, so that each block has the form

$$K(\lambda) = \begin{bmatrix} \lambda & & * \\ & \ddots & \\ 0 & & \lambda \end{bmatrix}$$

You will note that

$$K(\lambda) = \lambda I + N$$

where N is nilpotent, or $K(\lambda)$ is 1×1 .

- (3) "Jordanize" each triangular block. Assume that $K_1(\lambda)$ is $m \times m$, where $m > 1$. By construction $K_1(\lambda)$ pertains to an invariant subspace for which there is a unique vector x for which

$$N^{m-1}x \neq 0 \quad \text{and} \quad N^m x = 0.$$

Thus $N^{m-1}x$ is an eigenvector of $K_1(\lambda)$, the unique eigenvector. Define

$$y_i = N^{i-1}x \quad i = 1, 2, \dots, m.$$

Expand the set $\{y_i\}_{i=1}^m$ as a basis of $\mathbb{C}_m$. Define

$$S_1 = \begin{bmatrix} y_m & y_{m-1} & \cdots & y_1 \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}.$$

Then

$$NS_1 = \begin{bmatrix} 0 & y_m & y_{m-1} & \cdots & y_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix}.$$

So

$$S_1^{-1}NS_1 = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & 0 & \ddots \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix}.$$

We conclude that

$$S_1^{-1}K_1(\lambda)S_1 = \begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \ddots & \ddots \\ & & & & \ddots & 1 \\ 0 & & & & & \lambda \end{bmatrix}$$

- (4) Assemble all of the blocks to form the Jordan form. For example, the block $K_1(\lambda)$ and the corresponding similarity transformation S_1 studied above can be treated in the assembly process as follows: Define the $n \times n$ matrix

$$\hat{S}_1 = \begin{bmatrix} I & 0 & 0 \\ 0 & S_1 & 0 \\ 0 & 0 & I \end{bmatrix}$$

where S_1 is the block constructed above and placed in the $n \times n$ matrix in the position that $K_1(\lambda)$ was extracted from the block triangular form of A . Repeat this for each of the blocks pertaining to minimally invariant subspaces. This gives a sequence of block diagonal matrices $\hat{S}_1, \hat{S}_2, \dots, \hat{S}_k$. Define $T = \hat{S}_1\hat{S}_2\cdots\hat{S}_k$. It has the form

$$T = \begin{bmatrix} \hat{S}_1 & & & 0 \\ & \hat{S}_2 & & \\ & & \ddots & \\ 0 & & & \hat{S}_k \end{bmatrix}$$

Together with the original matrix P that transformed the matrix to the minimal invariant subspace blocked form and the unitary matrix

V used to triangularize A , it follows that

$$A = PVT \begin{bmatrix} J_{n_1}(\lambda_1) & & & \\ & \ddots & & \\ & & 0 & \\ & & & J_{n_k}(\lambda_k) \end{bmatrix} (PVT)^{-1} = SJS^{-1}$$

with $S = PVT$.

□

Example 8.3.1. Let

$$J = \begin{bmatrix} 2 & 1 & & & & \\ 0 & 2 & & & & \\ & & 2 & & & \\ & & & 3 & 1 & 0 \\ & & & 0 & 3 & 1 \\ & & & 0 & 0 & 3 \\ & & & & & -1 \end{bmatrix}$$

In this example, there are four blocks, with two of the blocks pertaining to the single eigenvalue 2. For the first block there is the single eigenvector e_1 , but the invariant subspace is $\mathfrak{S}(e_1, e_2)$. For the second block, the eigenvector, e_3 , generates the one dimensional invariant subspace. The block pertaining to the eigenvalue 3 has the single eigenvector e_4 while the minimal invariant subspace is $\mathfrak{S}(e_4, e_5, e_6)$. Finally, the one dimensional subspace pertaining to the eigenvalue -1 is spanned by e_7 . The minimal invariant polynomial is $q(\lambda) = (\lambda - 2)^2(\lambda - 3)^3(\lambda + 1)$.

8.4 Convergent matrices

Using the Jordan normal form, the study of convergent matrices becomes relatively straightforward and simpler.

Theorem 8.4.1. *If $A \in M_n$ and $\rho(A) < 1$. Then*

$$\lim_{k \rightarrow \infty} A^k = 0.$$

Proof. We assume A is a Jordan matrix. Each Jordan block $J_k(\lambda)$ can be written as

$$J_k(\lambda) = \lambda I_k + N_k$$

where

$$N_k = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix} \text{ is nilpotent.}$$

Now

$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{m_k}(\lambda_k) \end{pmatrix}.$$

We compute, for $m > n_k$

$$\begin{aligned} (J_{n_k}(\lambda_k))^m &= (\lambda I + N)^m \\ &= \lambda^m I + \sum_{j=0}^{m_k} \lambda^{m-j} N^j \binom{m}{j} \end{aligned}$$

because $N^j = 0$ for $j > n_k$. We have

$$\lambda^{m-j} \binom{m}{j} \rightarrow 0 \text{ as } m \rightarrow \infty$$

since $|\lambda| < 1$. The results follows. $\square$

8.5 Exercises

1. Prove Theorem 8.2.1.
2. Find a 3×3 matrix that has the same eigenvalues are the squares of the roots of the equation $\lambda^3 - 3\lambda^2 + 4\lambda - 5 = 0$.
3. Suppose that A is a square matrix with $\sigma(A) = \{3\}$, $m_a(3) = 6$, and $m_g(3) = 3$. Up to permutations of the blocks show all possible Jordan normal forms for A .

4. Let $A \in M_n(\mathbb{C})$ and let $x_1 \in \mathbb{C}^n$. Define $x_{i+1} = Ax_i$ for $i = 1, \dots, n-1$. Show that $V = \mathfrak{S}(\{x_1, \dots, x_n\})$ is an invariant subspace of A . Show that V contains an eigenvector of A .
5. Referring to the previous problem, let $A \in M_n(R)$ be a permutation matrix. (i) Find starting vectors so that $\dim V = n$. (ii) Find starting vectors so that $\dim V = 1$. (iii) Show that if $\lambda = 1$ is a simple eigenvalue of A then $\dim V = 1$ or $\dim V = n$.

Chapter 9

Hermitian and Symmetric Matrices

Example 9.0.1. Let $f: D \rightarrow R$, $D \subset R^n$. The *Hessian* is defined by

$$H(x) = h_{ij}(x) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \in M_n.$$

Since for functions $f \in C^2$ it is known that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

it follows that $H(x)$ is symmetric.

Definition 9.0.1. A function $f: R \rightarrow R$ is *convex* if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for $x, y \in D$ (domain) and $0 \leq \lambda \leq 1$.

Proposition 9.0.1. If $f \in C^2(D)$ and $f''(x) \geq 0$ on D then $f(x)$ is convex.

Proof. Because $f'' \geq 0$, this implies that $f'(x)$ is increasing. Therefore if $x < x_m < y$ we must have

$$f(x_m) \leq f(x) + f'(x_m)(x_m - x)$$

and

$$f(x_m) \leq f(y) + f'(x_m)(x_m - y)$$

Chapter 9

Hermitian and Symmetric Matrices

Example 9.0.1. Let $f: D \rightarrow R$, $D \subset R^n$. The *Hessian* is defined by

$$H(x) = h_{ij}(x) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \in M_n.$$

Since for functions $f \in C^2$ it is known that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

it follows that $H(x)$ is symmetric.

Definition 9.0.1. A function $f: R \rightarrow R$ is *convex* if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for $x, y \in D$ (domain) and $0 \leq \lambda \leq 1$.

Proposition 9.0.1. If $f \in C^2(D)$ and $f''(x) \geq 0$ on D then $f(x)$ is convex.

Proof. Because $f'' \geq 0$, this implies that $f'(x)$ is increasing. Therefore if $x < x_m < y$ we must have

$$f(x_m) \leq f(x) + f'(x_m)(x_m - x)$$

and

$$f(x_m) \leq f(y) + f'(x_m)(x_m - y)$$

by the Mean Value Theorem. Therefore,

$$\frac{y - x_m}{y - x} f(x_m) + \frac{x_m - x}{y - x} f(x_m) \leq \frac{y - x_m}{y - x} f(x) + \frac{x_m - x}{y - x} f(y)$$

or

$$f(x_m) \leq \frac{y - x_m}{y - x} f(x) + \frac{(x_m - x)}{y - x} f(y).$$

It is easy to see that

$$x_m = \frac{y - x_m}{y - x} x + \frac{x_m - x}{y - x} y.$$

Defining $\lambda = \frac{y - x_m}{y - x}$, it follows that $1 - \lambda = \frac{x_m - x}{y - x}$. □

Definition 9.0.2. We say that $f: D \rightarrow R$, where $D \subset R^n$ is convex, is a convex function if f is a convex function on every line in D .

Theorem 9.0.1. Suppose $f \in C^2(D)$ and $H(x)$ is positive definite. Then f is convex on D .

Proof. Let $x \in D$ and η be some direction. Then $x + \lambda\eta$ is a line in D . We compute $\frac{d^2}{d\lambda^2} f(x + \lambda\eta)$

$$\frac{d}{d\lambda} f = \nabla f \cdot \eta = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \cdot [\eta_1 \dots \eta_n].$$

Now

$$\frac{d}{d\lambda} \left(\frac{\partial f}{\partial x_1} \right) = \nabla \frac{\partial f}{\partial x_1} \cdot \eta = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} \\ \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} \end{bmatrix} \cdot [\eta_1, \dots, \eta_n].$$

Hence we see that

$$\frac{d^2}{d\lambda^2} f(x + \lambda\eta) = \eta^T H(x) \eta \geq 0$$

by assumption. □

Example 9.0.2. Let $A = [a_{ij}] \in M_n$. Consider the quadratic form on $\mathbb{C}^n$ or $\mathbb{R}^n$ defined by

$$\begin{aligned} Q(x) &= x^T A x = \sum a_{ij} x_j x_i \\ &= \frac{1}{2} \sum (a_{ij} + a_{ji}) x_j x_i \\ &= x^T \frac{1}{2} (A + A^T) x. \end{aligned}$$

Since the matrix $A + A^T$ is symmetric the study of quadratic forms is reduced to the symmetric case.

Example 9.0.3. Let $Lf = \sum_{i,j=1}^n a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}$. L is called a partial differential operator. By the combination of devices above (assuming $f \in C^2$ for example) we can study the symmetric and equivalent partial differential operator

$$Lf = \sum_{i,j=1}^n \frac{1}{2} (a_{ij} + a_{ji}) \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

In particular, if $A + A^T$ is positive definite the operator is called elliptic.

Other cases are

- (1) hyperbolic,
- (2) degenerate/parabolic.

Characterizations of Hermitian matrices. Recall

- (1) $A \in M_n$ is Hermitian if $A^* = A$.
- (2) $A \in M_n$ is called skew-Hermitian if $A = -A^*$.

Here are some facts

- (a) If A is Hermitian the diagonal is real.
- (b) If A is skew-Hermitian the diagonal is imaginary.
- (c) $A + A^*$, AA^* and A^*A are all Hermitian if $A \in M_n$.
- (d) If A is Hermitian then A^k , $k = 0, 1, \dots$, are Hermitian. A^{-1} is Hermitian if A is invertible.

- (e) $A - A^*$ is skew-Hermitian.
 (f) $A \in M_n$ yields the decomposition

$$A = \underbrace{\frac{1}{2}(A + A^*)}_{\text{Hermitian}} + \underbrace{\frac{1}{2}(A - A^*)}_{\text{Skew Hermitian}}$$

- (g) If A is Hermitian iA is skew-Hermitian. If A is skew-Hermitian then iA is Hermitian.

Theorem 9.0.2. *Let $A \in M_n$. Then $A = S + iT$ where S and T are Hermitian. Moreover this is unique.*

Proof.

$$\begin{aligned} A &= \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) \\ &= S + iT \end{aligned}$$

where $S = \frac{1}{2}(A + A^*)$ and

$$\begin{aligned} iT &= \frac{1}{2}(A - A^*) \\ \Rightarrow T &= -\frac{i}{2}(A - A^*). \quad \square \end{aligned}$$

Theorem 9.0.3. *Let $A \in M_n$ be Hermitian. Then*

- (a) x^*Ax is real for all $x \in \mathbb{C}^n$;
 (b) All the eigenvalues of A are real;
 (c) S^*AS is Hermitian for all $S \in M_n$.

Proof. For (a) we have

$$x^*Ax = \sum a_{ij}x_jx_i.$$

The conjugate is

$$\begin{aligned} \overline{x^*Ax} &= \sum \bar{a}_{ij}\bar{x}_jx_i = \sum a_{ji}\bar{x}_jx_i \\ &= \sum a_{ij}x_j\bar{x}_i + x^*Ax \quad \square \end{aligned}$$

Theorem 9.0.4. *Let $A \in M_n$. Then A is Hermitian if and only if at least one of the following holds:*

(a) $\langle Ax, x \rangle = x^* Ax$ is real $\forall x \in \mathbb{C}^n$.

(b) A is normal with real eigenvalues.

(c) S^*AS is Hermitian for all $S \in M_n$.

Proof. Hermitian $\Rightarrow$ (a), (b), or (c) are obvious. (a) $\Rightarrow$ Hermitian. Prove using e_j and e_k . We have

$$\langle Ae_j + e_k, e_j + e_k \rangle = a_{jj} + a_{kk} + a_{jk} + a_{kj}$$

$$\Rightarrow a_{jk} + a_{kj} \text{ is real} \Rightarrow \text{Im } a_{jk} = -\text{Im } a_{kj}.$$

Similarly

$$\langle Aie_j + e_k, ie_j + e_k \rangle = a_{jj} + a_{kk} + ia_{kj} - ia_{jk}$$

$$\Rightarrow i(a_{kj} - a_{jk}) \text{ is real} \Rightarrow \text{Re } a_{kj} = \text{Re } a_{jk}.$$

All this gives $A^* = A$.

(c) $\Rightarrow$ Hermitian S^*AS is Hermitian $\Rightarrow S^*AS$ is similar to a real diagonal matrix. Therefore A is similar to a real diagonal matrix. Just let $S = I$ to get A is Hermitian. $\square$

Theorem 9.0.5 (Spectral Theorem). *Let $A \in M_n$ be Hermitian. Then A is unitarily (similar) equivalent to a real diagonal matrix. If A is real Hermitian, then A is orthogonally similar to a real diagonal matrix.*

9.1 Variational Characterizations of Eigenvalues

Let $A \in M_n$ be Hermitian. Assume

$$\lambda_{\min} \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max}.$$

Theorem 9.1.1 (Rayleigh–Ritz). *Let $A \in M_n$, and let the eigenvalues of A be ordered as above. Then*

$$\lambda_{\max} = \lambda_n = \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

$$\lambda_{\min} = \lambda_1 = \min_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

Proof. Let $x_1 \dots x_n$ be the linearly independent and orthogonal eigenvectors of A . Any vector $x \in \mathbb{C}^n$ has the representation

$$x = \sum \alpha_j x_j.$$

Then $\langle Ax, x \rangle = \sum \alpha_j^2 \lambda_j$ and $\langle x, x \rangle = \sum \alpha_j^2$. Hence

$$\frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \frac{\sum_j \alpha_j^2 \lambda_j}{\sum_i \alpha_i^2}.$$

Since $\left\{ \frac{\alpha_j^2}{\sum \alpha_i^2} \right\}$ is nonnegative and sums to 1, it follows that

$$\frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

is a convex combination of the eigenvalues, whence the theorem follows. In particular take $x = x_n$ to achieve the maximum (minimum). $\square$

Remark 9.1.1. This result gives just the largest and smallest eigenvalues. How can we achieve the intermediate eigenvalues? We have already considered this problem somewhat in conjunction with the power method. In that consideration we employed the bi-orthogonal eigenvectors. For a Hermitian matrix, the families are the same. So we could characterize the eigenvalues in a manner similar to that discussed previously. However, the following characterization is simpler.

Theorem 9.1.2. *Let $A \in M_n$ be Hermitian with eigenvalues as above and corresponding eigenvectors $x_1 \dots x_n$. Then*

$$(*) \quad \lambda_{n-k} = \max_{\substack{x \perp \{x_n, \dots, x_{n-k+1}\} \\ x \neq 0}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

The following result is even more general

$$(*) \quad \lambda_{n-k} = \min_{\{w_n, w_{n-1}, \dots, w_{n-k+1}\}} \max_{\substack{x \perp \{w_n, w_{n-1}, \dots, w_{n-k+1}\} \\ x \neq 0}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

Proof. The Rayleigh–Ritz argument above gives $(*)$ directly. $\square$

9.2 Matrix inequalities

When the underlying matrix is symmetric or positive definite, certain powerful inequalities can be established. The first inequality is a consequence of the cofactor result Proposition 2.5.1.

Theorem 9.2.1. *Let $A \in M_n(\mathbb{C})$ be positive definite. Then $\det A \leq a_{11}a_{22} \cdots a_{nn}$*

Proof. Expanding in minors, the determinant of A

$$\det A = a_{11} \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & & a_{nn} \end{vmatrix} + \det \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix}$$

Since A is positive definite, it follows that each of its principal submatrices is also. Therefore, the first term in the right side of the equality above is positive, while the second term is negative by the cofactor result Proposition 2.5.1. To clarify, it is important to note that if A is Hermitian and invertible, so also is its inverse. In particular, if A is positive definite, we know its determinant is positive and so when we write

$$\det \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix} = - \sum a_{1j} a_{1k} a^{jk}$$

it is clear this term is negative. Therefore,

$$\det A \leq a_{11} \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & & a_{nn} \end{vmatrix}$$

whence the result follows inductively. $\square$

Corollary 9.2.1. *Let $S_1, S_2, \dots, S_r$ be any partition of the integers $\{1, \dots, n\}$, and let $A_1, A_2, \dots, A_r$ be the principal submatrices of A pertaining to the indices of each subdivision. Then*

$$\det A = \det A_1 \det A_2 \cdots \det A_r$$

An important consequence of Theorem 9.2.1 is one of the most famous determinantal inequalities. It is due to Hadamard.

Theorem 9.2.1. *Let B be an arbitrary nonsingular real square matrix. Then*

$$(\det B)^2 \leq \prod_{i=1}^n \sum_{j=1}^n |b_{ij}|^2$$

Proof. Define $A = B^T B$. Then

$$\text{diag} A = \left(\sum_{j=1}^n |b_{1j}|^2, \dots, \sum_{j=1}^n |b_{nj}|^2 \right)$$

whence the result follows from Theorem 9.2.1 since $\det A = (\det B)^2$. $\square$

9.3 Exercises

1. Prove Corollary 9.2.1.
2. Establish Theorem 9.2.1 in the case the matrix is complex.
3. Establish a result similar to Theorem 9.2.1 for rectangular matrices.

Chapter 10

Nonnegative Matrices

10.1 Definitions

Nonnegative matrices are simply those with all nonnegative entries. We will eventually distinguish various types of such matrices substantially on how they are nonnegative or conversely where they are zero. A particular type of nonnegative matrix, the type that has no off-diagonal block to be zero, will turn out to have the greatest value to us. If nonnegative matrices distinguished themselves in only minor ways from general matrices, they would not occupy the high position of importance they enjoy in the modern literature. Indeed, many remarkable properties of nonnegative matrices have been uncovered. Owing substantially to the many applications to economics, probability, and engineering, the subject has now for more than a century been one of the hottest areas of research within the subject.

Definition 10.1.1. $A \in M_n(R)$ is called **nonnegative** if $a_{ij} \geq 0$ for $1 \leq i, j \leq n$. We use the notation $A \geq 0$ for nonnegative matrices. If $a_{ij} > 0$ for all i, j we say A is *fully* (or *strictly*) *positive*. (Notation. $A \gg 0$.)

In this section we need some special notation. Let $x \in \mathbb{C}^n$, $x = (x_1, \dots, x_n)^T$. Then

$$|x| = (|x_1|, \dots, |x_n|)^T.$$

We say that $x \in R^n$ is *positive* if $x_i \geq 0$, $1 \leq i \leq n$. We say that $x \in R^n$ is *fully* (or *strictly*) *positive* if $x_i > 0$, $1 \leq i \leq n$. We use the notation $x \geq 0$ and $x \gg 0$ for positive and strictly positive vectors, respectively.

Note that if $A \geq 0$ we have $\|A\|_\infty = \|Ae\|_\infty$ where $e = (1, 1, \dots, 1)^T$.

To give an idea of the direction we are heading, let us suppose that $A \in M_n(R)$ is a nonnegative matrix. Let $\lambda \neq 0$ be an eigenvalue of A with pertaining eigenvector $x \geq 0$. Thus $Ax = \lambda x$. (We will prove that this comes to pass for every nonnegative square matrix.) Suppose further that the vector Ax is zero on the index set S . That is $(Ax)_i = 0$ if $i \in S$. Since $\lambda \neq 0$ it follows that $x_i = 0$ if $i \in S$. Let S^C be the complement of S in the integers $\{1, \dots, n\}$, and let P and $P^\perp$ be the projections to S and S^C , respectively. So, $Px = 0$ and $P^\perp x = x$. Also, it is easy to see that $I_n = P + P^\perp$, and hence

$$PAP^\perp x = \lambda Px = 0$$

Therefore $PAP^\perp = 0$. When we write

$$A = (P + P^\perp) A (P + P^\perp)$$

in block form

$$A = (P + P^\perp) A (P + P^\perp) = \begin{bmatrix} PAP & PAP^\perp \\ P^\perp AP & P^\perp AP^\perp \end{bmatrix}$$

we see that A has the special form

$$A = \begin{bmatrix} PAP & 0 \\ P^\perp AP & P^\perp AP^\perp \end{bmatrix}$$

This particular calculation reveals in a simple way the consequences of zero components to eigenvectors of nonnegative matrices. Zero components of eigenvectors implies zero blocks of A . It would be correct to observe that this argument requires a nonnegative eigenvector. Nonetheless, there are still some very special properties of the remaining eigenvalues of A , particularly those with the modulus equal to the spectral radius. It will be convenient to have a name for the index set where a nonnegative vector $x \in R^n$ is strictly positive.

Definition 10.1.2. Let $x \in R^n$ be nonnegative, $x \geq 0$. Let S be the index set such that $x_i \neq 0$ if $i \in S$. We call S the *support* of the vector x .

10.2 General Theory

Definition 10.2.1. For $A \in M_n(R)$ and $\lambda \notin \sigma(A)$ we define the *resolvent* of A by

$$R(\lambda) = (\lambda I - A)^{-1}.$$

This function behaves much like a rational function in complex variable theory. It has poles at each $\lambda \in \sigma(A)$. The order of the pole is the same as the order of the zero of λ in the minimal polynomial of A .

Theorem 10.2.1. *Suppose $A \in M_n(R)$ is nonnegative and $\lambda > \rho(A)$, then $R(\lambda) = (\lambda I - A)^{-1} \geq 0$.*

Proof. Apply the Neumann series

$$(\star) \quad (\lambda I - A)^{-1} = \sum_{k=0}^{\infty} \lambda^{-(k+1)} A^k.$$

Since $\lambda > 0$ and $A \geq 0$, it is apparent that $R(\lambda) \geq 0$. □

Remark 10.2.1. It is apparent that $(\star)$ holds upon noting that

$$(\lambda I - A)^{-1} \left(1 - \left(\frac{A}{\lambda} \right)^{n+1} \right) = \sum_0^n \lambda^{-(k+1)} A^k$$

and that $|\lambda| > \rho(A)$ yields convergence.

Theorem 10.2.2. *Suppose $A \in M_n(R)$ is nonnegative. Then the spectral radius $\rho(A)$ of A is an eigenvalue of A with at least one eigenvector, $x \geq 0$.*

Proof. Suppose for each $y \geq 0$, $R(\lambda)y$ remains bounded as $\lambda \downarrow \rho$. If $x \in \mathbb{C}^n$ is arbitrary

$$\begin{aligned} |R(\lambda)x| &= \left| \sum_{n=0}^{\infty} \lambda^{-(n+1)} A^n x \right| \\ &\leq \sum_{n=0}^{\infty} |\lambda|^{-(n+1)} A^n |x| \\ &= R(|\lambda|)|x| \end{aligned}$$

for all $\lambda, |\lambda| > \rho$. It follows that $R(\lambda)x$ is uniformly bounded in the region $|\lambda| > \rho$, and this is impossible.

Now let $y_0 \geq 0$ be a vector for which $R(\lambda)y_0$ is unbounded as $\lambda \downarrow \rho$, and let $\| \cdot \|$ denote a vector norm. For $\lambda > \rho$, set

$$z(\lambda) = R(\lambda)y_0 / \|R(\lambda)y_0\|,$$

where $\|R(\lambda)y_0\| \uparrow \infty$. Now $z(\lambda) \subset \{x \mid \|x\| = 1\}$, and the latter is compact. Therefore $\{z(\lambda)\}_{\lambda > \rho}$ has a cluster point x_0 with $x_0 \geq 0$ and $\|x_0\| = 1$. Since

$$(\rho I - A)z(\lambda) = (\rho - \lambda)z(\lambda) + y_0 / \|R(\lambda)y_0\|$$

and since $\lambda \rightarrow \rho$ and $\|R(\lambda)y_0\| \uparrow \infty$ we have

$$\lim_{\lambda \rightarrow \rho} (\rho I - A)z(\lambda) = \lim_{\lambda \rightarrow \rho} (\rho I - A)x_0 = 0. \quad \square$$

Corollary 10.2.1. *Suppose $A \in M_n(R)$ is nonnegative and $Az = \lambda z$ for some $z \gg 0$, then $\lambda = \rho(A)$.*

Proof. Since $\rho(A) \subset \sigma(A^T)$ there is a vector $y \geq 0$ for which $A^T y = \rho y$. If $\lambda \neq \rho$, we must have $\langle z, y \rangle = 0$. But $z \gg 0$ and this is impossible. $\square$

Corollary 10.2.2. *Suppose $A \in M_n(R)$ is strictly positive and $z \geq 0$ such that $Az = \rho z$, then $z \gg 0$.*

Proof. If $z_i = 0$, then $\sum_{j=1}^n a_{ij}z_j = 0$, and thus $z_j = 0$ for all j since $a_{ij} > 0$, for all j . Hence $z = 0$, a contradiction. $\square$

This result will be strengthened in the next section by another result that has the same conclusion but a much weaker hypotheses.

Theorem 10.2.3. *Suppose $A \in M_n(R)$ is nonnegative with spectral radius $\rho(A)$ and define*

$$\begin{aligned} r_m &= \min_i \sum_j a_{ij} & r_M &= \max_i \sum_j a_{ij} \\ c_m &= \min_j \sum_i a_{ij} & c_M &= \max_j \sum_i a_{ij} \end{aligned}$$

then both inequalities below hold.

$$\begin{aligned} r_m &\leq \rho(A) \leq r_M \\ c_m &\leq \rho(A) \leq c_M \end{aligned}$$

Moreover, if $A, B \in M_n(R)$ then $\rho(A + B) \geq \max(\rho(A), \rho(B))$.

Proof. We prove the second set of inequalities. Let $x \in R_n$ be the eigenvector pertaining to $\rho(A)$. Since $x \geq 0$ we can normalize x so that $\sum x_i = 1$. Then

$$\sum_{j=1}^n a_{ij}x_j = \rho(A)x_i, \quad i = 1, \dots, n$$

Sum both side in i to obtain

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_j &= \sum_{i=1}^n \rho(A) x_i \\ \sum_{j=1}^n \left(\sum_{i=1}^n a_{ij} \right) x_j &= \rho(A)\end{aligned}$$

Replacing $\sum_{i=1}^n a_{ij}$ by c_M makes the sum on the left larger, that is

$$\rho(A) \leq \sum_{j=1}^n \left(\sum_{i=1}^n a_{ij} \right) x_j \leq c_M \sum_{i=1}^n x_i = c_M$$

Similarly replacing $\sum_{i=1}^n a_{ij}$ by c_m makes the sum on the left smaller, and therefore $\rho(A) \geq c_m$. Thus, $c_m \leq \rho(A) \leq c_M$. The other inequality, $r_m \leq \rho(A) \leq r_M$, can be easily established by applying what has just been proved to the matrix A^T noting of course that $\rho(A) = \rho(A^T)$.

The proof that $\rho(A+B) \geq \max(\rho(A), \rho(B))$ is an easy consequence of a Neumann series argument. $\square$

We could just as well prove the first inequality and apply it to the transpose to obtain the second. This proof is given below.

Alternative Proof. Let x be the eigenvector pertaining to $\rho(A)$ and $x_k = \max_i x_i$. Then

$$\rho x_k = \sum_{j=1}^n a_{kj} x_j \leq \left(\sum_{j=1}^n a_{kj} \right) \max x_j.$$

Hence

$$\rho \leq \sum_{j=1}^n a_{kj} \leq \max_k \sum_{j=1}^n a_{kj} = t.$$

To obtain $s \leq \rho$, let $y \geq 0$ satisfy

$$A^T y = \rho y$$

and suppose $\|y\|_1 = 1$. Then we have on the one hand

$$\langle e, A^T y \rangle = \rho \langle e, y \rangle = \rho$$

and on the other hand applying convexity

$$\begin{aligned}\langle e, A^T y \rangle &= \sum_i \sum_j a_{ij}^T y_j \\ &= \sum_j y_j \left(\sum_i a_{ji} \right) \geq \min_j \sum_i a_{ji} = s\end{aligned}$$

□

Corollary 10.2.1. (i) If $\rho(A)$ is equal to one of the quantities c_m or c_M , then all of the sums $\sum_i a_{ij}$ are equal for j in the support of the eigenvector x pertaining to ρ .

(ii) If $\rho(A)$ is equal to one of the quantities r_m or r_M , then all of the sums $\sum_j a_{ij}$ are equal for i in the support of the eigenvector y of A^T pertaining to ρ . (iii) If all the row sums (resp. column sums) are equal, the spectral radius is equal to this value.

Proof. Assume that $\rho(A) = c_M$. Let S denote the support (recall Definition 10.1.2) of the eigenvector x . Assume also that the eigenvector x is normalized so that $\sum_{j=1}^n x_j = \sum_{j \in S} x_j = 1$. Following the proof of Theorem 10.2.3 we have

$$\sum_{j=1}^n \left(\sum_{i=1}^n a_{ij} \right) x_j = \sum_{j \in S} \left(\sum_{i=1}^n a_{ij} \right) x_j = c_M$$

Since $\frac{1}{c_M} (\sum_{i=1}^n a_{ij}) \leq 1$ for each $j = 1, \dots, n$ it follows that

$$\sum_{j \in S} \frac{1}{c_M} \left(\sum_{i=1}^n a_{ij} \right) x_j \leq \sum_{j \in S} x_j$$

Clearly if for some $j \in S$ we have $\frac{1}{c_M} (\sum_{i=1}^n a_{ij}) < 1$ the inequality above will become a strict inequality, and the result is proved. The result is proved similarly for c_m

(ii) Apply the proof of (i) to A^T . □

10.3 Mean Ergodic Theorem

Let $A \in M_n(R)$. We have seen that understanding the nature of powers A^k , $k = 1, 2, \dots$ leads to interesting conclusions. For example if $\lim_{k \rightarrow \infty} A^k = P$,

it is easy to see that P is idempotent ($P^2 = P$) and $AP = PA = P$. It is therefore a projection to the *fixed space* of A . A less restrictive property is *mean convergence*:

$$M_k = k^{-1}(I + A + A^2 + \cdots + A^k)$$

$$\lim_{k \rightarrow \infty} M_k = P.$$

Note that if $\rho(A) < 1$ we have $A^k \rightarrow 0$. Hence $M_k \rightarrow 0$. On the other hand if $\rho(A) > 1$ the M_k become unbounded. Hence mean convergence has value only when $\rho(A) = 1$.

Theorem 10.3.1 (Mean Ergodic Theorem for matrices). *Let $A \in M_n(R)$ with $\rho(A) = 1$. If $\{A^k\}_{k=1}^\infty$ is bounded then $\lim_{k \rightarrow \infty} M_k = P$, where P is a projection, commuting with A , onto the fixed space of A .*

Proof. We need a norm $\|\cdot\|$, vector and matrix. We have $\|A^k\| < C$ for $k = 1, 2, \dots$. It is easy to see that

$$(\star) \quad \begin{aligned} AM_k - M_k &= \frac{1}{k} \left(\sum_0^k A^{j+1} \right) - \frac{1}{k} \left(\sum_0^k A^j \right) \\ &= \frac{1}{k} (I + A^{k+1}) \\ &\leq \frac{1}{k} (1 + C). \end{aligned}$$

Since the matrices M_k are bounded they have a cluster point P . This means there is a subsequence M_{k_j} that converges to P . If there is another cluster point Q (i.e. there is a subsequence $M_{\ell_j} \rightarrow Q$), then we compare P and Q . First we know

$$\begin{aligned} \|M_{k_j} - P\| &< \frac{\varepsilon}{2(1+C)} \\ \|M_{\ell_j} - Q\| &< \frac{\varepsilon}{2(1+C)}. \end{aligned}$$

From $(\star)$ we have $AP = P$ and $AQ = Q$, whence $M_{k_j}P = P$ and $M_{\ell_j}Q = Q$ for all $k = 1, 2, \dots$. Hence

$$P - Q = M_{\ell_j}(M_{k_j} - P) - M_{k_j}(M_{\ell_j} - Q)$$

or

$$\|P - Q\| \leq \varepsilon.$$

Thus $P = Q$ and M_k converges to some matrix P . We have $AP = PA = P$ hence $M_k P = P$ for all k and therefore $P^2 = P$. Clearly the range of P consists of all fixed vectors under A . $\square$

Corollary 10.3.1. *Let $A \in M_n(R)$ and suppose that $\lim_{k \rightarrow \infty} A^k = P$, then $\lim_{k \rightarrow \infty} M_k$ exists and equals P .*

Definition 10.3.1. Let $A \in M_n(R)$ have $\rho(A) = 1$. Then A is said to be *mean ergodic* if

$$\lim k^{-1}(I + A + \cdots + A^k)$$

exists.

Clearly, if $A \in M_n(R)$ satisfies $\rho(A) = 1$ and A^k is bounded, then A is mean ergodic. (Previous theorem.) Conversely, if A is mean ergodic then the sequence A^k is bounded. This can be proved by resorting to the Jordan canonical form and showing A^k is bounded if and only if A_J^k is bounded, where A_J is the Jordan canonical form of A .

Lemma 10.3.1. *Let $A \in M_n(R)$ with $\rho(A) = 1$. Suppose $\lambda = 1$ is a simple pole of the resolvent, and the only eigenvalue of modulus 1. Then $A = P + B$ where $P = \lim_{\lambda \rightarrow 1} (\lambda - 1)R(\lambda)$ is a projection onto the fixed space of A , $PB = BP = 0$ and $\rho(B) < 1$.*

Proof. We know that P exists and using the Neumann series it is clear that $AP = PA = P$

$$\begin{aligned} A \lim_{\lambda \rightarrow 1} (\lambda - 1)(\lambda I - A)^{-1} &= A(\lambda - 1) \sum_{k=0}^{\infty} \lambda^{-(k+1)} A^k \\ &= \lim_{\lambda \rightarrow 1} (\lambda - 1) \sum_{k=0}^{\infty} \lambda^{-(k+1)} A^{k+1} \\ &= \lim_{\lambda \rightarrow 1} (\lambda - 1) \lambda \left[\sum_{k=0}^{\infty} \lambda^{-(k+1)} A^k - \lambda^{-1} \right] \\ &= \lim_{\lambda \rightarrow 1} (\lambda - 1) R(\lambda) = P. \end{aligned}$$

That is, $AP = P$. The other assertions follow similarly. Also, it follows that

$$P^2 = \lim_{\lambda \rightarrow 1} P R(\lambda) = P$$

which is to say that P is a projection. We have, as well, that $Ax = x \Rightarrow Px = x$, again using the Neumann series.

Now define $B = A - P$. Then $BP = PB = 0$. Suppose $Bx = \alpha x$, where $|\alpha| \geq 1$. Then $Px = 0$, for

$$\alpha Px = PBx = 0.$$

Therefore $Ax = \alpha x$. Here $\alpha \neq 1$ implies $x = 0$ by hypothesis and $\alpha = 1$ implies $Px = x$, or $x = 0$, also. This contradicts the assumption that $|\alpha| \geq 1$. $\square$

Theorem 10.3.2. *Let $A \in M_n(R)$ satisfy $A \geq 0$ and $\rho(A) = 1$. The following are equivalent:*

(a) $\lambda = 1$ is a simple pole of $R(\lambda)$.

(b) $\{A^k\}_1^\infty$ is bounded.

(c) A is mean ergodic.

Moreover

$$\lim_{\lambda \rightarrow 1} (\lambda - 1)R(\lambda) = \lim_{k \rightarrow \infty} \frac{1}{k}(I + A + \cdots + A^k)$$

if either limit exists.

10.4 Irreducible Matrices

Irreducible matrices form one of the cornerstones of positive matrix theory. Because the condition of irreducibility is both natural and often satisfied, their applications are far and wide.

Definition 10.4.1. Let $A \in M_n(\mathbb{C})$. The **zero pattern** of A is the set of ordered pairs $S = \{(i, j) \mid a_{ij} = 0\}$.

Example 10.4.1. For example with

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 3 & 0 \end{bmatrix}$$

$$S = \{(1, 1), (2, 2), (2, 3), (3, 3)\}.$$

Definition 10.4.2. A generalized permutation matrix B is any matrix having the same zero pattern as a permutation matrix.

This means of course that a generalized permutation matrix has exactly one nonzero entry in each row and one nonzero entry in each column. Generalized permutation matrices are those nonnegative matrices with nonnegative inverses.

Theorem 10.4.1. *Let $A \in M_n(R)$ be invertible and $A \geq 0$. Then A is invertible with nonnegative inverse if and only if A is a generalized permutation matrix.*

Proof. First, if A is a generalized permutation matrix, define the matrix B by

$$b_{ij} = \begin{cases} 0 & \text{if } a_{ij} = 0 \\ \frac{1}{a_{ij}} & \text{if } a_{ij} \neq 0 \end{cases}$$

Then $A^{-1} = B^T \geq 0$. On the other hand suppose $A^{-1} \geq 0$ and A has on some row two non zero entries, say a_{i,j_1} and a_{i,j_2} . Since $A^{-1} \geq 0$ is invertible there is a nonzero entry in the i^{th} column, say $a_{ki}^{-1} > 0$. Now compute multiply $A^{-1}A$. It is easy to see that

$$(A^{-1}A)_{k,j_1} > 0 \quad \text{and} \quad (A^{-1}A)_{k,j_2} > 0$$

which contradicts that $A^{-1}A = I$. □

Example 10.4.1. The analysis for 2×2 matrices can be carried out directly. Suppose that

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

There are two cases: (1) $\det A > 0$. In this case we conclude that $b = c = 0$. Thus A is a generalized permutation matrix. (2) $\det A < 0$. In this case we conclude that $a = d = 0$. Again A is a generalized permutation matrix. As these are the only two cases, the result is verified.

Definition 10.4.3. Let $A, B \in M_n(R)$. We say that A is **cogredient** to B if there is a permutation matrix P such that

$$B = P^T A P$$

Note that two cogredient matrices are similar, indeed they are unitarily similar for the very special class of unitary matrices given by permutations. It is important to note that since AP merely interchanges the columns of A and $P^T AP$ then interchanges the rows of AP . From this we observe that both A and $B = P^T AP$ have exactly the same elements, though permuted.

Definition 10.4.4. Let $A \in M_n(R)$. We say that A is **irreducible** if it is not cogredient to a matrix of the form

$$\left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array} \right]$$

where the block matrices A_1 and A_4 are square matrices. Otherwise the matrix is called **reducible**.

Example 10.4.2. All diagonal matrices are reducible.

There is another way to describe irreducible (and reducible) matrices in terms of projections. Let $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ and define P_S to be the projection to the coordinate directions by

$$P_S e_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases}$$

Furthermore define

$$P_S^\perp = I - P_S$$

Then P_S and $P_S^\perp$ are orthogonal projections. That is, the product $P_S P_S^\perp = 0$ and by definition $P_S + P_S^\perp = I$. If S^C denotes the complement of S in $\{1, 2, \dots, n\}$, it is easy to see that $P_S^\perp = P_{S^C}$.

Proposition 10.4.1. Let $A \in M_n(R)$. Then A is irreducible if and only if there is no set of indices $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ for which $P_S A P_S^\perp = 0$.

Proof. Suppose there is a set of indices $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ for which $P_S A P_S^\perp = 0$. Define the permutation matrix as from the permutation that takes the first k integers $1, \dots, k$ to the integers $j_1, \dots, j_k$ and the integers $k+1, \dots, n$ to the complement S^C . Then

$$P^T A P = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array} \right]$$

The entries in the block A_2 are those with rows given from the rows corresponding to the indices in S and columns corresponding to the indices in S^C . Thus A is reducible. Conversely, suppose that A is reducible and that P is a permutation matrix such that

$$P^T A P = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array} \right]$$

For definiteness, let us assume that A_1 is $k \times k$ and A_4 is $(n-k) \times (n-k)$. Define $S = \{j_i \mid p_{j_i, i} = 1, i = 1, \dots, k\}$. and $T = \{j_i \mid p_{j_i, i} = 1, i = k+1, \dots, n\}$. Because P is a permutation, it follows that $T = S^C$ and $P_S A P_S^\perp = 0$, which proves the converse. $\square$

As we know for a nonnegative matrix $A \in M_n(R)$ the spectral radius is an eigenvalue of A with pertaining eigenvector $x \geq 0$. When the additional assumption of irreducibility of A is added the conclusion can be strengthened to $x \gg 0$. To prove this result we establish a simple result from which the $x \gg 0$ follows almost directly.

Theorem 10.4.2. *Let $A \in M_n(R)$ be nonnegative and irreducible. Suppose that $y \in R_n$ is nonnegative with exactly $1 \leq k \leq n-1$ nonzero entries. Then $(I + A)y$ has strictly more nonzero entries.*

Proof. Suppose the nonzero entries of y are $S = \{j_1, \dots, j_k\}$. Since $(I + A)y = y + Ay$, it follows immediately that $((I + A)y)_{j_i} \neq 0$ for $j_i \in S$, and therefore there are at least as many nonzero entries in $(I + A)y$ as there are in y . In order that the number of nonzero entries not increase, we must have for each index $i \notin S$ that $(Ay)_i = 0$. With P_S defined as the projection to the standard coordinates with indices from S we conclude therefore that $P_S^\perp A P = 0$, which means that A is reducible, a contradiction. $\square$

Corollary 10.4.1. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if $(I + A)^{n-1} > 0$.*

Proof. Suppose that A is irreducible and $y \geq 0$ is any vector in R_n . Then $(I + A)y$ has strictly more nonzero coordinates and thus $(I + A)^2 y$ has even more nonzero coordinates. We see that the number of nonzero coordinates of $(I + A)^k y$ must increase by at least one until the maximum of n nonzero coordinates is reached. This must occur by the $(n-1)^{\text{th}}$ power. Now apply this to the vectors of the standard basis $e_1, \dots, e_n$. For example, $(I + A)^{n-1} e_k \gg 0$. This means that the k^{th} column of $(I + A)^{n-1}$ is strictly positive. The result follows.

Suppose conversely that $(I + A)^{n-1} > 0$ but that A is reducible. Then there is a permutation matrix P such that $P^T A P$ has the form

$$P^T A P = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array} \right]$$

It is easy to see that all the powers of A have the same form — with respect to the same blocks. Also, $(I + A)^{n-1} = I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i$. So

$$\begin{aligned} P^T (I + A)^{n-1} P &= P^T \left(I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i \right) P \\ &= I + \sum_{i=1}^{n-1} \binom{n-1}{i} P^T A^i P^T \end{aligned}$$

has the same form

$$P^T (I + A)^{n-1} P = \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline \tilde{A}_3 & \tilde{A}_4 \end{array} \right]$$

and this proves the result. $\square$

Corollary 10.4.2. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if for each pair of indices (i, j) there a positive power $1 \leq k \leq n$ such that $(A^k)_{ij} > 0$.*

Proof. Suppose that A is irreducible. Since $(I + A)^{n-1} = I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i > 0$, it follows that $A(I + A)^{n-1} = A + \sum_{i=1}^{n-1} \binom{n-1}{i} A^{i+1} > 0$. Thus for each pair of indices (i, j) , $(A^k)_{ij} > 0$ for some k .

On the other hand suppose that A is reducible. Then there is a permutation matrix P for which

$$P^T A P = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array} \right]$$

Follow the same argument as above to establish that $A(I + A)^{n-1}$ must have the same form as $P^T A P$ with the same zero pattern. Thus there is no positive power $1 \leq k \leq n$ such that $(A^k)_{ij} > 0$ for any of the pairs of indices corresponding to the (upper right) zero block above. $\square$

Another characterization of irreducibility arises when we have $Ax \leq ax$ for some nonzero vector $x \geq 0$. This type of domination-type condition appears to be quite weak. Nonetheless, it is equivalent to the others.

Corollary 10.4.3. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if $Ax \leq ax$ for some nonzero vector $x \geq 0$, then $x \gg 0$.*

Proof. If $x_i = 0$, then $a_{ij} = 0$ whenever $x_j \neq 0$. Define $S = \{1 \leq j \leq n \mid x_j \neq 0\}$. Then with P_S as the orthogonal projection to the standard vectors e_i , $i \in S$, it must follow that $P_S A P_S^\perp = 0$. Since S is strictly contained in $\{1, 2, \dots, n\}$, it follows that A is reducible.

Now suppose that A is reducible, which means we can assume it has the form

$$A = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array} \right]$$

For definiteness, let us assume that A_1 is $k \times k$ and A_4 is $(n-k) \times (n-k)$ and of course $k \geq 1$. Let $v \in R_{n-k}$ be the nonzero nonnegative vector which satisfies $A_4 v = \rho(A_4) v$. Let $u = 0 \in R_k$. Define the vector $x = u \oplus v$. Then

$$(Ax)_i = \begin{cases} (A_1 u)_i = 0 = \rho(A_4) u_i & \text{if } 1 \leq i \leq k \\ (A_1 v)_i = \rho(A_4) v_i & \text{if } k+1 \leq i \leq n \end{cases}$$

The conditions of that the hypothesis are met without $x \gg 0$. □

Corollary 10.4.4 (Subinvariance). *Let $A \in M_n(R)$ be nonnegative and irreducible with spectral radius ρ . Suppose that there is a positive number s and nonnegative vector $z \geq 0$ for which*

$$(Az)_i \leq sz_i, \quad \text{for } i = 1, 2, \dots, n$$

Then (i) $z \gg 0$, and (ii) $\rho \leq s$. If in addition $\rho = s$, then $Az = \rho z$.

Proof. Clearly, the same inequality holds for powers of A . For $A(Az) \leq Az \leq sz$. Now apply induction. Next suppose that $z_i = 0$. By Corollary 10.4.2 we must have for each index pair i, j that $(A^k)_{ij} > 0$ for some positive integer k making $(A^k z)_i \leq sz_i$ impossible for some k . Thus $z \gg 0$. To prove the second assertion, let x be the eigenvector of A^T pertaining ρ . Then

$$\langle Az, x \rangle = \langle z, A^T x \rangle = \rho \langle z, x \rangle \leq s \langle z, x \rangle \quad (1)$$

Since $x \gg 0$, the innerproduct $\langle z, x \rangle > 0$. Hence $\rho \leq s$.

Finally, if $\rho = s$ and $(Az)_i < \rho z_i$ for some index i then we conclude from (1) that $\rho < \rho$, a contradiction. □

Theorem 10.4.3. *Let $A \in M_n(R)$ be nonnegative and irreducible. If $x \geq 0$ is the eigenvector pertaining to $\rho(A)$, i.e. $Ax = \rho(A)x$, then $x \gg 0$.*

Proof. We have that $(I + A)$ is also nonnegative and irreducible, with spectral radius $1 + \rho(A)$ and pertaining eigenvector x . Since

$$(I + A)^{n-1}x > (1 + \rho(A))^{n-1}x \gg 0$$

it follows that $x \gg 0$. □

Corollary 10.4.5. *(i) If $\rho(A)$ is equal to one of the quantities c_m or c_M from Theorem 10.2.3, then all of the sums $\sum_i a_{ij}$ are equal for $j = 1, \dots, n$.*

(ii) If $\rho(A)$ is equal to one of the quantities r_m or r_M from Theorem 10.2.3, then all of the sums $\sum_j a_{ij}$ are equal for $i = 1, \dots, n$. (iii) Conversely, if all the row sums (resp. column sums) are equal, the spectral radius is equal to this value.

Proof. Both are direct consequences of Corollary 10.2.1 and Theorem 10.4.3 which imply that the eigenvectors (of A and A^T , resp.) pertaining to ρ are strictly positive. □

10.4.1 Sharper estimates of the maximal eigenvalue

Sharper estimates for the maximal eigenvalue (spectral radius) are possible. The following results assembles much of what we have considered above. We begin with a basic inequality that has an interesting geometric interpretation.

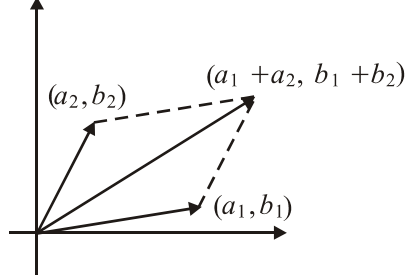
Lemma 10.4.1. *Let $a_i, b_i, i = 1, \dots, n$ be two positive sequences. Then*

$$\min_j \frac{b_j}{a_j} \leq \frac{\sum_{j=1}^n b_j}{\sum_{j=1}^n a_j} \leq \max_j \frac{b_j}{a_j}$$

Proof. We proceed by induction. The result is trivial for $n = 1$. For $n = 2$ the result is a simple consequence of vector addition as follows. Consider the ordered pairs (a_i, b_i) , $i = 1, 2$. Their sum is $(a_1 + a_2, b_1 + b_2)$. It is a simple matter to see that the slope of this vector must lie between the slopes of the summands, which is to say

$$\min_j \frac{b_j}{a_j} \leq \frac{\sum_{j=1}^2 b_j}{\sum_{j=1}^2 a_j} \leq \max_j \frac{b_j}{a_j}$$

as is shown below.



Now assume by induction the result holds up to $n - 1$. Then we must have that

$$\begin{aligned} \frac{\sum_{j=1}^n b_j}{\sum_{j=1}^n a_j} &= \frac{\sum_{j=1}^{n-1} b_j + b_n}{\sum_{j=1}^{n-1} a_j + a_n} \leq \max \left[\frac{\sum_{j=1}^{n-1} b_j}{\sum_{j=1}^{n-1} a_j}, \frac{b_n}{a_n} \right] \\ &\leq \max \left[\max_{1 \leq j \leq n-1} \frac{b_j}{a_j}, \frac{b_n}{a_n} \right] = \max_{1 \leq j \leq n} \frac{b_j}{a_j} \end{aligned}$$

A similar argument serves to establish the other inequality. $\square$

Theorem 10.4.4. *Let $A \in M_n(R)$ be nonnegative with row sums r_k and spectral radius ρ . Then*

$$\min_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \leq \rho \leq \max_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \quad (2)$$

Proof. First assume that A is irreducible. Let $x \in R^n$ be the principal nonnegative eigenvector of A^T , so $A^T x = \rho x$ and $x \gg 0$. Assume also that x has been normalized so that $\sum x_i = 1$. By summing both sides of $A^T x = \rho x$, we see that $\sum r_k x_k = \rho$. Since ρ^2 is the spectral radius of A^2 , we also have that $(A^T)^2 x = (A^2)^T x = \rho^2 x$. Summing both sides, it follows that $\sum_{k=1}^n \sum_{j=1}^n r_j a_{jk}^T x_k = \rho^2$. Now

$$\rho = \frac{\rho^2}{\rho} = \frac{\sum_{k=1}^n \sum_{j=1}^n r_j a_{jk}^T x_k}{r_k x_k} = \max_k \frac{\sum_{j=1}^n a_{kj} r_j}{r_k}$$

by Lemma 10.4.1. The reverse inequality is proved similarly. In the case that A is not irreducible, we take the limit $A_k \downarrow A$ where the A_k are irreducible matrices. The result holds for each A_k , and the quantities in the inequality are continuous in the limiting process. Some small care must be taken in the case that A has a zero row. $\square$

Of course, we have not yet established that the estimates (2) are sharper than the basic inequality $\min_k r_k \leq \rho \leq \max_k r_k$. That is the content of following result, whose proof is left as an exercise.

Corollary 10.4.6. *Let $A \in M_n(R)$ be nonnegative with row sums r_k . Then*

$$\min_k r_k \leq \min_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \leq \rho \leq \max_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \leq \max_k r_k$$

We can continue this kind of argument even further. Let $A \in M_n(R)$ be nonnegative with row sums r_i and let x be the nonnegative eigenvector pertaining to ρ . Then $(A^T)^3 x = \rho^3 x$ or expanded

$$\sum_{m,k,j} a_{ij}^T a_{jk}^T a_{km}^T x_m = \rho^3 x_i$$

Summing both sides yields

$$\sum_{m,k,j} r_j a_{jk}^T a_{km}^T x_m = \rho^3$$

Therefore,

$$\begin{aligned} \rho &= \frac{\rho^3}{\rho^2} = \frac{\sum_m \sum_k \sum_j r_j a_{jk}^T a_{km}^T x_m}{\sum_m \sum_j r_j a_{jm}^T x_m} \\ &\leq \max_m \frac{\sum_k \sum_j r_j a_{jk}^T a_{km}^T}{\sum_j r_j a_{jm}^T} = \max_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \end{aligned}$$

by Lemma 10.4.1, thus yielding another estimate for the spectral radius. The estimate is two-sided as those above. This is summarized in the following result.

Theorem 10.4.5. *Let $A \in M_n(R)$ be nonnegative with row sums r_i . Then*

$$\min_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \leq \rho \leq \max_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \quad (3)$$

Remember the complete proof as developed above does require the assumption irreducibility and the passage of the limit. Let's consider an example and see what these estimates provide.

Example 10.4.3. Let

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 5 & 3 & 5 \\ 1 & 1 & 4 \end{bmatrix}$$

The eigenvalues of A are -2 , 2 , and 8 . Thus $\rho = 8$. The row sums are $r_1 = 7$, $r_2 = 13$, $r_3 = 6$, which yields the estimates $6 \leq \rho \leq 13$. The estimates from (2) are $\frac{64}{7}$, 8 , and $\frac{22}{3}$ giving the estimates

$$7.33333333 \leq \rho \leq 9.142857143$$

Finally, from (3) the estimates for ρ are

$$7.818181818 \leq \rho \leq 8.192307692$$

It remains to show that the estimates in (3) furnish an improvement to the estimates in (2). This is furnished by the following corollary, which is left as an exercise.

Corollary 10.4.7. *Let $A \in M_n(R)$ be nonnegative with row sums r_i . Then for each $m = 1, \dots, n$*

$$\min_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \leq \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \leq \max_k \left[\frac{1}{r_k} \sum_{j=1}^n a_{kj} r_j \right] \quad (4)$$

These results are by no means the end of the story on the important subject of eigenvalue estimation.

10.5 Stochastic Matrices

Definition 10.5.1. A matrix $A \in M_n(R)$, $A \geq 0$, is called (row) *stochastic* if each row sum of A is 1 or equivalently $Ae = e$. (Recall, $e = (1, 1, \dots, 1)^T$.) Similarly, a matrix $A \in M_n(R)$, $A \geq 0$, is called column stochastic if A^T is stochastic.

A direct consequence of Corollary 10.2.1(iii) is that the spectral radius of stochastic matrices must be one.

Theorem 10.5.1. *Let $A \in M_n(R)$ row or column stochastic. Then the spectral radius of A is $\rho(A) = 1$.*

Lemma 10.5.1. *Let $A, B \in M_n(R)$ be nonnegative row stochastic matrices. Then for any number $0 \leq c \leq 1$, the matrices $cA + (1 - c)B$ and AB are also row stochastic. The same result holds for column stochastic matrices.*

Theorem 10.5.2. *Every stochastic matrix is mean ergodic, and the peripheral spectrum is fully cyclic.*

Proof. If $A \geq 0$ is stochastic, so also is A^k , $k = 1, 2, \dots$. Therefore A^k is bounded. Also $\rho(A) \leq \max_i \sum_k a_{ij} = 1$. Hence the conditions of the previous theorems are met. This proves A is mean ergodic. $\square$

Stochastic matrices—as a set—have some interesting properties. Let $S_n \subset M_n(R)$ of stochastic matrices. Then S_n is a convex set, it is also closed. Whenever a closed convex set is determined, the characterization of its extreme points is desired.

Recall, the (matrix) point $A \in S_n$ is called an *extreme point* if whenever

$$A = \sum \lambda_j A_j$$

$\sum \lambda_j = 1$, $\lambda_j \geq 0$ and $A_j \in S_n$ it follows that one of the A_j 's equals A and all λ_j but one are 0.

Examples. We can also view $S_n \subset R^{n^2}$. It is a convex subset of R^{n^2} and is moreover the intersection of the positive orthant of R^{n^2} with the n hyperplanes defined by $\sum_j a_{ij} = 1$, $i = 1, \dots, n$. [Note that we are viewing a vector $x \in R^{n^2}$ as

$$x = (\alpha_{11}, \dots, \alpha_{1n}, \alpha_{21}, \dots, \alpha_{2n}, \dots, \alpha_{n1}, \dots, \alpha_{nn})^T]$$

S_n is therefore a convex polyhedron in R^{n^2} . The dimension of S_n is $n^2 - n$.

Theorem 10.5.3. *$A \in S_n$ is an extreme point of S_n if and only if A has exactly one 1 in each row, and all other entries are zero.*

Proof. Let $C = c_{ij}$ be a (0,1)-matrix. (That is a matrix $c_{ij} = \begin{cases} 1 \\ 0 \end{cases}$ or for all $1 \leq i, j \leq n$.) Suppose that $C \in S_n$ then there is a unique j for which $c_{ij} = 1$. If $C = \lambda A + (1 - \lambda)B$ it is easy to see that $a_{1j} = b_{1j} = 1$ and moreover that $a_{1k} = b_{1k} = 0$ for all other $k \neq j$. It follows that C is an

extreme point. Conversely suppose $C \in S_n$ is *not* a $(0,1)$ matrix. Fix one row. Let

$$A_j = \begin{bmatrix} e_j & \\ c_2 & \rightarrow \\ \vdots & \\ c_n & \rightarrow \end{bmatrix}$$

where $e_j = (0, 0, \dots, 1, 0 \dots 0)$ and $c_2 \dots c_n$ are the respective rows of C .
 $j^{\text{th}} \text{ position}$

Then

$$C = \sum_1^n C_{1j} A_j.$$

If C_1 is not a $(0,1)$ row we are finished, since C cannot be an extreme point. Otherwise select any non $(0,1)$ row and repeat this argument with respect to that row. $\square$

Corollary 10.5.1. *Permutation matrices are extreme points.*

Definition 10.5.2. A lattice homomorphism of $\mathbb{C}^n$ is a linear map $A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ such that $|Ax| = A|x|$, for all $x \in \mathbb{C}^n$. A is a lattice isomorphism if both A and A^{-1} are lattice homomorphisms.

Theorem 10.5.4. $A \in S_n$ is a lattice homomorphism if and only if A is an extreme point of S_n . $A \in S_n$ is a lattice isomorphism if and only if A is a permutation matrix.

Theorem 10.5.5. $A \in S_n$ is a permutation matrix if and only if $\sigma(A) \subset \Gamma$.

$\Gamma = \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$. **Exercise.**

1. Show that any row stochastic matrix (i.e. nonnegative entries with row sums are equal 1) with at least one column having equal entries is singular.
2. Show that the only nonnegative matrices with nonnegative inverses must be diagonal matrices.
3. Adapt the argument from Section 10.1 to prove that the nonnegative eigenvector x pertaining to the spectral radius of an irreducible matrix must be strictly positive.

4. Suppose that $A, B \in M_n(R)$ are nonnegative matrices. Prove or disprove
 - (a) If A is irreducible then A^T is irreducible.
 - (b) If A is irreducible then A^p is irreducible, for every positive power $p \geq 1$.
 - (c) If A^2 is irreducible then A is irreducible.
 - (d) If A, B are irreducible, then AB is irreducible.
 - (e) If A, B are irreducible, then $aA + bB$ is irreducible for all non-negative constants $a, b \geq 0$.
5. Suppose that $A, B \in M_n(R)$ are nonnegative matrices. Prove that if A is irreducible, then $aA + bB$ is irreducible for all nonnegative constants $a, b > 0$.
6. Suppose that $A \in M_n(R)$ is nonnegative and irreducible and that the trace of A is positive, $\text{tr} A > 0$. Prove that $A^k > 0$ for some sufficiently large integer k .
7. Prove Corollary 10.2.1(ii) for $\rho(A) = r_m$.
8. Let $A \in M_n(R)$ be nonnegative with row sums $r_i(A)$ and column sums $c_i(A)$. Show that $\sum_{i=1}^n r_i(A^2) = \sum_{i=1}^n c_i(A) r_i(A)$.
9. Prove Corollary 10.4.6.
10. Prove Corollary 10.4.7. (Hint. Use Lemma 10.4.1.)
11. Let p be any positive integer and suppose that $A \in M_n(R)$ is nonnegative with row sums r_i . Define $r = (r_1, \dots, r_n)^T$. Recall that $(A^p r)_m$ refers to the m^{th} coordinate of the vector $A^p r$. Show that

$$\min_m \frac{(A^p r)_m}{(A^{p-1} r)_m} \leq \rho \leq \max_m \frac{(A^p r)_m}{(A^{p-1} r)_m}$$

(Hint. Assume first that A is irreducible.)

12. Use the estimates (2) and (3) to estimate the maximal eigenvalue of

$$A = \begin{bmatrix} 5 & 1 & 4 \\ 3 & 1 & 5 \\ 1 & 2 & 2 \end{bmatrix}$$

The maximal eigenvalue is approximately 7.640246936.

13. Use the estimates (2) and (3) to estimate the maximal eigenvalue of

$$A = \begin{bmatrix} 2 & 1 & 4 & 4 \\ 5 & 7 & 3 & 2 \\ 5 & 0 & 6 & 0 \\ 5 & 5 & 5 & 7 \end{bmatrix}$$

The maximal eigenvalue is approximately 14.34259731.

14. Suppose that $A \in M_n(R)$ is nonnegative and irreducible with spectral radius ρ , and suppose $x \in R^n$ is any strictly positive vector, $x \gg 0$. Define $\tau = \max_i \frac{(Ax)_i}{x_i}$. Show that $\rho \leq \tau$.

- 15.

$$\begin{aligned} A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \sigma(A) = \{1\} \quad C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma(B) = \{-1, 1\} \quad \rho_C(\lambda) = \lambda^3 + 1 = 0 \\ \lambda = -1, e^{i\pi/3}, e^{-i\pi/3} \end{aligned}$$