Math 211 Linear Algebra Skeleton Notes S. Waner

1. Row Operations

Definitions 1.1 A **field** is a set F with two binary operations $+, \times$, such that:

1. *Addition and multiplication are commutative:*

$\forall x, y \in F, x + y = y + x$ and $xy = yx$.

2. *Addition and multiplication are associative:*

$\forall x, y, z \in F, x + (y + z) = (x + y) + z$ and $x(yz) = (xy)z$..

3. *Addition and multiplication are unital:*

There exists a unique elements $0 \in F$ and $1 \in F$ (with $0 \neq 1$) such that $\forall x \in F, 0 + x = x + 0 = x$ and $1 \cdot x = x \cdot 1 = x$.

4. *Existence of inverses:*

For each element $x \in F$, there exists a unique element $(-x) \in F$ such that $x + (-x) = (-x) + x = 0$.

For each nonzero element $x \in F$, there exists a unique element $x^{-1} \in F$ such that $xx^{-1} = x^{-1}x = 1$.

5. *Multiplication distributes over addition:*

$\forall x, y, z \in F$, one has $x(y + z) = xy + xz$.

A **subfield** K of a field F is a subset that is a field in its own right under the operations inherited from F. We often refer to the elements of a field as **scalars**.

Examples 1.2

A. C, the field of complex numbers

B. R, the field of real numbers is a subfield of C.

C. The set Q of rational numbers, is a subfield of R and hence C.

D. Z_p, the field of integers modulo the prime p.

E. N (the set of non-negative integers) is not a subfield of R; nor is the set Z of integers.

Note Henceforth, we shall assume F to be a subfield of C.

Definitions 1.3 A **system of m linear equations in n unknowns** is a collection of equations of the form

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = y_1 \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = y_2 \\ &\cdots \\ &a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = y_m \ , \end{aligned}$$

where each $a_{ij} \in F$. The x_i are called the **unknowns.** The system is **homogeneous** if each $y_i = 0$. A **solution** of a system of linear equations is an n-tuple $(s_1, s_2, \ldots, s_n)$ of elements of F which satisfies the above system of equations.

Definitions 1.4 An $m \times n$ **matrix over the field** F is a function A from the set $\{(i, j) \mid 1 \le i \le m,\ 1 \le j \le n\}$ to F. We write $A(i, j)$ as A_{ij} or a_{ij}, and represent A by the following array of elements of F.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

We represent the system of equations in Definition 1.3 by the formal equation

$$AX = B,$$

where A is as above, $X = [x_1 \quad x_2 \quad \dots \quad x_n]^t$, and $B = [b_1 \quad b_2 \quad \dots \quad b_m]^t$.

Definition 1.5 An **elementary row operation** on the $m \times n$ matrix A over F is a rule of one of the following types:

1. The operation $R_i \to cR_i$ for $c \neq 0$ in F.

2. The operation $R_i \to R_i + cR_j$

3. The operation $R_i \leftrightarrow R_j$

The $m \times n$ matrices A and B are said to be **row equivalent** if B can be obtained from A by a finite sequence of elementary row operations.

Proposition 1.6
(a) Row equivalence is an equivalence relation.
(b) If A and B are row equivalent, then the set of solutions to $AX = 0$ ad $BX = 0$ coincide.

Note The converse to 1.6(b) is also true—see later.

Definition 1.7 The $m \times n$ matrix A is in **row reduced echelon form** if:

(a) The leading entry (i.e., first nonzero entry) of each row is a 1.

(b) The column of each leading entry contains only one nonzero entry (which must therefor be the leading entry itself).

(c) The leading entry of each row must be to the left of those of the rows below it, with rows of zeros (if any) at the bottom.

Examples in class.

Theorem 1.7
Every $m \times n$ matrix A is row-equivalent to a matrix in row reduced echelon form.

Notes

1. In a row reduced matrix, there can be no more than one leading entry per row. Also, since the column of each leading entry is clear, there can be no more than one leading entry per column.

2. If A is a square matrix in row reduced echelon form, then every column has a leading entry iff A is the identity matrix.
3. If A is row reduced echelon with leading entries in columns $i_1, i_2, \ldots, i_r$, and no leading entries in columns $j_1, j_2, \ldots, j_s$, then the system of equations $AX = 0$ has the form

$$\begin{aligned} x_{i_1} &= F_1(x_{j_1}, x_{j_2}, \ldots, x_{j_s}) \\ x_{i_2} &= F_2(x_{j_1}, x_{j_2}, \ldots, x_{j_s}) \\ &\ldots \\ x_{i_r} &= F_r(x_{j_1}, x_{j_2}, \ldots, x_{j_s}), \end{aligned}$$

where the F_k are linear combinations. (The rows of zeros don't say anything about the solution.) Thus, we can choose the x_{j_t} to be arbitrary, and still obtain a solution. In other words, there are infinitely many solutions.

Proposition 1.8
Let A be an $m \times n$ matrix.
(a) If $m < n$, then the system $AX = 0$ has infinitely many solutions.
(b) If $m = n$, then the system $AX = 0$ has only the zero the solution iff A is row-equivalent to the identity matrix.

Proof
(a) Since the solution set of the system $AX = 0$ is invariant under row operations, we can assume that A is in row reduced echelon form. Since $m < n$, A has fewer rows than columns, so there are not enough leading entries for all the columns (see the above note). Part (a) now follows from Note 3 above.
(b) If $m = n$, and A is row equivalent to the identity matrix, then $AX = 0$ has only the single solution 0. (Read off the solution, which does not depend on row operations.) Conversely, if $AX = 0$ has only the single solution 0, then row reduction must lead to one leading entry per column, or else there is a column without a leading entry, and we are in part (a), leading to more than one solution. Hence, row reduction must lead to the identity. ◆

Definition 1.9 The **augmented** matrix of a system $AX = B$ is the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}.$$

Note Row operations of the augmented matrix do not effect the solution of $AX = B$. Thus, to solve such a system, all one need do is row reduce the augmented matrix.

Proof of the following proposition is left as homework.

Proposition 1.10
(a) In the system $AX = B$, exactly one of the following can occur: a unique solution, no solutions, or infinitely many solutions.
(b) If $m < n$, then the system $AX = B$ either has no solutions, or has infinitely many solutions.

Exercise Set 1
1. Find all possible solutions of the following systems

(a) $\begin{aligned} x_1 + x_2 &= 1 \\ ix_1 - x_3 &= 0 \\ (2+i)x_2 + x_3 &= 0 \end{aligned}$ (b) $\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ ix_1 - 4x_2 &= 0 \\ (1-i)x_1 + 5x_2 - x_3 &= 0 \end{aligned}$

2. Prove that the interchange of two rows in a matrix may be accomplished by a sequence of elementary row operations of the other two types.
3. Prove Proposition 1.10.
4. Prove that, if A and B have entries in the subfield E of F, then all solutions of $AX = B$ are in E.

2. Matrix Algebra

Definitions 2.1 We define the **sum** of two matrices, **scalar multiples** of a matrix, and **product** of two matrices as follows. (All matrices are over the field F.)

(a) If A and B are both $m \times n$ matrices, then $A+B$ is the matrix over F whose entries are given by

$$[A+B]_{ij} = A_{ij} + B_{ij}.$$

(b) If A is a $m \times n$ matrix and $c \in F$, then cA is the matrix over F whose entries are given by

$$[cA]_{ij} = cA_{ij}.$$

(c) If A is an $m \times n$ matrix, and B is an $n \times p$ matrix, then AB is the matrix over F whose entries are given by

$$[AB]_{ij} = \sum_{k=0}^{n} A_{ik}B_{kj} \ .$$

We also define: $-A$ to be $(-1)A$, the zero matrix 0 to be the matrix all of whose entries are 0, and the identity matrix I to be the square matrix (one for each n) whose ijth entry is

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

Examples 2.2
A. Some numerical examples, including identity matrix.
B. The notation for a system of equations is consistent with the above definitions.
C. Example where $AB \neq BA$.

Proposition 2.3
The following properties hold for matrices A, B and C, and scalars λ, μ, whenever the expressions make sense.

(a) $(A+B)+C = A+(B+C)$	$(AB)C = A(BC)$		(associativity)
(b) $A+B = B+A$	—		(commutativity of addition)
(c) $A+0 = 0+A = A$	$AI = IA = A$		(identity)
(d) $A+(-A) = (-A)+A = 0$	*discussion to come*		(additive inversion)
(e) $\lambda(A+B) = \lambda A + \lambda B$	$(\lambda+\mu)A = \lambda A + \mu A$		(distributivity of sc. mult.)
(f) $(\lambda\mu)A = \lambda(\mu A)$	$(\lambda A)B = \lambda(AB)$		(associativity of sc. mult.)
(g) $1.A = A$			(identity for sc. mult.)
(h) $A(B+C) = AB + AC$	$(A+B)C = AC + BC$		(distributivity)

Note Numerous other properties, such as $0A = 0$ (whether the 0 on the left is a scalar of a matrix), $(-1)A = -A$, $-(A+B) = -A - B$, $A^nA^m = A^{n+m}$, etc., follow formally from those above. In fact, some of the properties above follow from others!

Definition 2.4 An **elementary matrix** is an $n\times n$ matrix E obtained from the identity matrix by performing a single row operation e on it. Thus, $E = e(I)$, and there are three types (illustration in class).

Proposition 2.5
(a) If A is any $n\times k$ matrix, and E is any $n\times n$ elementary matrix, then $EA = e(A)$.
(b) A is row-equivalent to B iff $A = PB$, where P is a product of elementary matrices.
(c) A is row-equivalent to I iff A is a product of elementary matrices.

Note Proposition 2.5 says that elementary row operations are nothing more than multiplication by elementary matrices.

Definition 2.6 If A is a $n\times n$ matrix, then A is **invertible** if there is an $n\times n$ matrix B such that $AB = BA = I$. When this occurs, we call B (the) **inverse of A** and write $B = A^{-1}$.

Examples
A. All elementary matrices.
B. If A is any row-reduced $n\times n$ matrix, then A is invertible iff A contains no rows of zeros (because how can you get a "1" by multiplying with a row of zeros?)
C. Method for calculating inverses. (Justification in the exercises.)

Proposition 2.6
Let A be any invertible $n\times n$ matrix over F. Then:
(a) A^{-1} is unique.
(b) A^{-1} is invertible, with inverse A.
(c) If B is another $n\times n$ invertible matrix, then AB is also invertible, with $(AB)^{-1} = B^{-1}A^{-1}$.
(d) If B is another $n\times n$ matrix, then AB is invertible iff B is invertible.

Corollary 2.7
(a) Arbitrary products of invertible matrices are invertible.
(b) Products of elementary matrices are invertible.
(c) If a matrix is row equivalent to I, then it is invertible.
(d) A and B are row equivalent iff $A = PB$, where P is some invertible matrix. (Compare 2.5(b).)

Proof of (c) in Exercise Set 2.

In fact, we have the following.

Theorem 2.8 (Invertibility)
If A is an $n \times n$ matrix, then the following are equivalent:
(a) A is invertible.
(b) A is row-equivalent to I.
(c) A is a product of elementary matrices.
(d) The system of equations $AX = 0$ has only the trivial solution.
(e) The system of equations $AX = Y$ has solutions for every choice of Y.
(f) A has a right inverse.
(g) A has a left inverse.

Proof
(a)⇒(b) Let C be the row-reduced form of A. Then $C = PA$, where P is a product of elementary matrices. P is invertible by Corollary 2.7(b). By Proposition 2.6(c), C is invertible, since A and P are. Thus, C must be the identity. (See Example B above.)
(b)⇒(c) Proposition 2.5(c)
(c)⇒(d) Write $A = E_1E_2 \ldots E_k$, where the E_i are elementary matrices. Then A is invertible, by Corollary 2.7(b) and so $AX = 0 \Rightarrow X = A^{-1}0 = 0$.
(d)⇒(e) If $AX = Y$ failed to have a solution for some Y, then A would have to reduce to a matrix with a row of zeros (or else the row operations that reduce A to the identity with also produce a solution). But then A would not be invertible, by Example B above.
(e)⇒(f) If we take Y_i to be the i th column of the identity matrix, an X_i to be a solution to $AX = Y_i$, then $AB = I$, where B is the matrix whose columns are the X_i.
(f)⇒(g) Write $AB = I$. Claim that, in fact, A must in fact be invertible. Otherwise, by Example B, A would reduce to a matrix S with a row of zeros, giving $PA = S$ where P is invertible and S has a row of zeros. But then $P = PI = PAB = SB$, which too has a row of zeros (look at the zero row of S times anything). But how an invertible matrix have a row of zeros?
(g)⇒(a) Write $BA = I$ for some B. Then B has a right inverse and is therefore invertible, by the argument above. Thus, A is also invertible, by Proposition 2.6(d). ◆

Exercise Set 2
H&K # 1, 2, 3, 7., p. 26 # 3, 9, 11
Also: Justify the method for calculating the inverse in Example C. That is, show that it works.
Hand In:

1. Prove Proposition 2.3 (b) and (h) (only one of them).
2. Prove that if $0A = 0$, (where the 0's denote zero matrices).
3. Prove that, if A and B are $n\times n$ matrices, then AB is invertible iff A and B are both invertible. Conclude that, if a product $A_1A_2 \ldots A_n$ of $n\times n$ matrices is invertible, then so is each of the factors A_i.
4. (H&K, p. 21 # 4) Let $A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix}$. Find elementary matrices $E_1, E2, \ldots, E_k$ such that $A = E_1E2 \ldots E_k$.
5. Show that, if A and B are two square matrices such that $AB = 0$, then either A or B is not invertible.

3. Vector Spaces and Subspaces

Let F be a field.

Definition 3.1 A **Vector space V over F** is a set V of **vectors** together with a operations $V\times V\longrightarrow V$, called **vector addition** (we write $v + w$ for the sum of v and w in V) and $F\times V\longrightarrow V$, called **scalar multiplication** (we write cv for the product of $c \in F$ and $v \in V$) such that the following hold.
(a) *V is an abelian group under addition.*
Addition is commutative: $v + w = w + v$ for all $v, w \in V$.
Addition is associative: $v + (w + u) = (v + w) + u$ for all $v, w, u \in V$
Addition is unital: there is a unique element $0 \in V$ such that $v + 0 = 0 + v = v$ for all $v \in V$.
Existence of additive inverses: for every $v \in V$, there exists a unique element $-v \in V$ such that $v + (-v) = (-v) + v = 0$.
(b) *Scalar multiplication is unital, associative and distributive.*
Unital: $1v = v$ for all $v \in V$.
Associative: $c(dv) = (cd)v$ for all $v \in V$ and $c, d \in F$.
Distributive: $c(v+w) = cv + cw$ and $(c+d)v = cv + dv$ for all $v, w \in V$ and $c, d \in F$.

Examples 3.2
A. $\{0\}$ is a vector space over F, called the **trivial vector space.**
B. R is a v.s. over R. **C.** R^n is a v.s. over R.
D. F^n is a v.s. over F. (Ex. Set 3) **E.** $F^{m\times n}$, the set of $m\times n$ matrices over F.
F. $F[x]$ is a v.s. over F. **G.** Map(S, F) where S is any set. (Ex. Set 3)
H. $Q[\sqrt{2}\,]$ is a v.s. over Q. **I.** C is a v.s. over R.
J. R is a v.s. over Q. More generally,
K. Any field extension E of F is a v.s. over F.

Lemma 3.3 Let V be a v.s. over F. Then,
(a) $0.v = 0$ for all $v \in V$.
(b) $c.0 = 0$ for all $c \in F$.
(b) $(-c)v = c(-v) = -(cv)$ for all $v \in V$ and $c \in F$.

Proof in Exercises

Definition 3.4 Let V be a v.s. over F, and let $\{v_1, v_2, \ldots, v_n\}$ be a collection of vectors in V. Then a **linear combination of the v_i** is a vector of the form

$$v = c_1v_1 + c_2v_2 + \ldots + c_nv_n \text{ ,} = \sum_{i=1}^{n} c_iv_i \text{ .}$$

where $c_1, c_2, \ldots, c_n$ are elements of F.

Example 3.5
A. Every vector in R^3 is a linear combination of **i**, **j**, and **k**.
B. Every vector in F^n is a linear combination of $e_1, e_2, \ldots, e_n$, where e_i is the n-tuple with 0 everywhere except for a 1 in the ith place.
C. If the $m \times n$ matrix A is row-equivalent to B, then, as elements of F^n, every row of B is a linear combination of the rows in A.

Definition 3.6 A **subspace** of the vector space V over F is a subset W of V so that W inherits the structure of a vector space over F from V. (In other words, W is itself a vector space over F under the operations of V.) When W is a subspace of V, we shall write $V < W$.

Proposition 3.7 (Test for a Subspace)
The non-empty subset W of V is a subspace iff, for all $v, w \in W$ and $c \in F$, one has $cv + w \in F$.

Examples 3.8
A. The zero subspace **B.** $Q < R < C$
C. $H(n)$, the set of $n \times n$ **Hermitian matrices** ($A_{ij} = \bar{A}_{ji}$) is a subspace of $M(n, C) = C^{n \times n}$ over R, but not over C. (Look at multiplication by i.)
D. The solution set of a system of homogeneous linear equations—this follows from matrix algebra.: $A(cX + Y) = c(AX) + AY$.
E. The intersection of an arbitrary collection of subspaces is a subspace.

We obtain more examples from the following.

Definition 3.9 Let V be a v.s. over F, and let $\{v_\alpha : \alpha \in A\}$ be any collection of vectors. Then the **span** of $\{v_\alpha\}$ is the set $\langle v_\alpha \rangle$ of all *finite* linear combinations of the v_α. We say that$\{v_\alpha\}$ **spans V** if $\langle v_\alpha \rangle = V$.

Proposition 3.9 (Interpretation of the Span)
If $\{v_\alpha\}$ is any collection of vectors, then $\langle v_\alpha \rangle$ is the intersection of all subspaces W that contain $\{v_\alpha\}$. (In other words, it is the "smallest subspace containing the v_α.")

Examples 3.10

A. $\langle e_i \rangle = \mathrm{R}^n$ **B.** span indep. of row operations.
C. Finding the span of some vectors in R^4 (in class)
D. Finding the span of some vectors in C^4 (in class)
E. $\{1, x, x^2, \ldots, x^n, \ldots\}$ spans $F[x]$.

Definition 3.11 The **sum** of the subspaces $W_1, W_2, \ldots, W_k$ is given by
$W_1+W_2+ \ldots + W_k = \langle \{w_1+w_2+ \ldots +w_k \mid w_i \in W_i\} \rangle$.
Thus it is the smallest subspace of V containing all the W_i.

Example 3.12 The sum of any two distinct lines through the origin in R^3 is a plane.

Exercise Set 3
1. Prove Lemma 3.3.
2. (Do not hand in) Hoffman & Kunze, p. 33, #1 (verify that F^n is a v.s. over F.)
3. Verify that Map(S, V) is a v.s. over F (see Example 3.2 G).
4. H&K, p. 33, #3, 7.
5. (Do not hand in) H&K, p. 39 #1, 2, 4, 6.
6. (H&K p. 40 #9) Prove that if $V = W_1+W_2$ with $W_1 \cap W_2 = \{0\}$, then each vector v in V can be written *uniquely* in the form $v = r + s$ with $r \in W_1$ and $s \in W_2$.

4. Bases and Dimension

Definition 4.1 A set S of vectors is called **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_r$ and vectors $v_1, v_2, \ldots, v_r \in S$ such that $c_1v_1 + c_2v_2 + \ldots + c_rv_r = 0$. A set of vectors that is not linearly dependent is called **linearly independent.**

Notes
1. S is linearly independent iff no vector in S can be expressed as a linear combination of (finitely many of) the others.
2. Any subset of a linearly independent set is also linearly independent.
3. Any set containing the zero vector is linearly dependent.
4. S is linearly independent iff every finite subset of S is linearly independent.

Lemma 4.2 (Row Operations and Independence)
If B is row equivalent to A, then the rows of the matrix A are independent iff the rows of B are independent.

Proof Note that every row of B is a linear combination of the rows of A. Thus, if the rows of B were dependent, it would follow that the rows of A would also be dependent. *Mutatis mutandis*, If the rows of A are dependent, so are the rows of B. ◆

Examples 4.3

A. Vectors in F^n; row reduction test.

B. The vectors $e_1, e_2, \ldots, e_n \in F^n$.

Continuing Theorem 2.8, we have the following.

Proposition 4.4 (More Criteria for Invertibility)
Let A be an $n \times n$ matrix. Then the following are equivalent.
(a) A is invertible.
(b) The rows of A are linearly independent.
(c) The columns of A are linearly independent.

Proof

(a)⇒(b) If A is invertible, then it is row equivalent to he identity, whose rows are independent. Thus the rows of A must be independent, by the Lemma.

(b)⇒(c) If the rows of A are linearly independent, and B is the row reduced echelon equivalent of A, then the rows of B must also be independent, so there can be no rows of zeros. But then, since B is square and row reduced with no rows of zeros, it must be the identity. Since now A is equivalent to the identity, it is invertible, and hence a product of elementary matrices by Theorem 2.8. This also implies that it is *column* equivalent to the identity, whence its columns are also independent, by the Lemma.

(c)⇒(a) By the same argument as the first part of (b)⇒(c), if the columns of A are linearly independent, it must be invertible. ◆

Definition 4.5 A **basis** of the vector space V over F is a linearly independent generating set $\mathcal{B}$. The v.s. V over F is called **finite dimensional** if it has a finite basis. Otherwise, it is **infinite dimensional**.

Examples 4.6

A. The **standard basis** $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ of F^n.

B. With $V = F[x]$, take $\mathcal{B} = \{1, x, x^2, \ldots, x^n, \ldots\}$.

C. F^n is finite dimensional over F

D. $F[x]$ is not finite dimensional over F. (Exercises)

E. R is infinite dimensional over Q.

Theorem 4.7 (Equivalent Definitions of a Basis—"Basis Theorem")
The following are equivalent:
(a) $\mathcal{B}$ is a basis for V.
(b) $\mathcal{B}$ is a linearly independent generating set.
(c) $\mathcal{B}$ is a maximal linearly independent set of vectors in V.
(d) $\mathcal{B}$ is a minimal generating set of V.

Proof

(a)⇒(b) This is Definition 4.5.

(b)⇒(c) If $\mathcal{B}$ were not a maximal linearly independent set, then it would be a subset of some other linearly independent set $\mathcal{C}$. But, with $c \in \mathcal{C} - \mathcal{B}$, one has c a linear combination

of the elements in $\mathcal{B}$, since $\mathcal{B}$ is a generating set. But this means that $\mathcal{C}$ is not an independent set. #

(c)⇒(d) First, we show $\mathcal{B}$ is a generating set, so let $v \in V$, and s'pose that $v \notin \langle\mathcal{B}\rangle$. Then claim that $\mathcal{B}\cup\{v\}$ would be independent, contradicting (b). Indeed, if it is *dependent*, then there must be a linear combination equal to zero with the coefficient of v non-zero. But then we contradict the fact that $v \notin \langle\mathcal{B}\rangle$. Next, $\mathcal{B}$ must be a *minimal* generating set, or else it would not be independent.

(d)⇒(a) We need only prove linear independence. But if it were not, then one of its elements would be a linear combination of the others, thus permitting us to throw it out and obtain a smaller generating set. # ◆

We now digress a little to prove an astounding fact.

Definition 4.A A **partial ordering** is a relation $\leq$ that is reflexive, transitive and antisymmetric ($a\leq b$ and $b\leq a \Rightarrow a=b$). A **partially ordered set** (poset) is a set with a partial ordering. If every two elements are comparable (ie., either $a \leq b$, $b \leq a$ or both), then we have an **ordering**.

Examples 4.B

1. Z, R **2.** The set of subsets of any set S.

Definition 4.C A **chain** in a poset is a collection of elements so that every two are comparable. In other words, it is an ordered subset of a poset.

Example 4.D

Look at some chains in the set of subsets of S.

Definition 4.E If $\mathcal{A}$ is a subset of the poset $\mathcal{P}$, then an **upper bound** of $\mathcal{A}$ is an element $u \in \mathcal{P}$ such that $u \geq a$ for every $a \in \mathcal{A}$. A **maximal** element of the subset $\mathcal{A}$ of $\mathcal{P}$ is an element $a \in \mathcal{A}$ (←—note) such that $a \leq b$, $a \neq b \Rightarrow b \notin \mathcal{A}$.

Lemma 4.F (Zorn)
If $\mathcal{P}$ is a partially ordered set such that every chain in $\mathcal{P}$ has an upper bound in $\mathcal{P}$, then $\mathcal{P}$ has at least one maximal element.

Theorem 4.G
Every vector space has a basis.

Proof Let V be a vector space, take $\mathcal{F}$ to be the poset of linearly independent subsets of V, and let $\mathcal{C}$ be a chain of linearly independent sets. It suffices to show that $\mathcal{C}$ has an upper bound in $\mathcal{F}$. But if $\mathcal{B}$ is the union of the subsets in $\mathcal{C}$, then we claim that $\mathcal{B}$ is independent. Indeed, any finite subset of $\mathcal{B}$ must live in some member of $\mathcal{C}$, and thus be independent. The claim now follows from Note 4 after Definition 4.1. ◆

Corollary 4.H
R has a basis as a vector space over Q. (No one has found one yet!)

Back to the main material.

Theorem 4.8
Let V be finite dimensional.
(a) Every finite generating set of V contains a basis.
(b) Every finite linearly independent set of V can be enlarged to a basis.
(c) Any two bases for V have the same number of elements. (In particular, every basis is finite!)

Proof
(a) Start with a finite generating set and keep removing vectors that are linear combinations of others. Eventually, you hit a minimal set (since the original set is finite, so you can't go on forever…) But Theorem 4.7 says that such a minimal set must be a basis.
(b) Start with the linearly indep. set $\mathcal{A}$ and a finite generating set $\mathcal{G}$ and keep adding vectors from $\mathcal{G}$ to $\mathcal{A}$, not bothering with those that are already in the span of what you have. At each stage, you still have a linearly independent set, so you ultimately wind up with a spanning independent set—i.e., a basis.
(c) If one has more than the other, express each element of the bigger basis $\{b_i\}$ as a linear combination of elements of the smaller. (If there are infinitely many, just use a large number of them.) The associated coefficient matrix must reduce to a matrix with at least one row of zeros (since it has too many rows). But this means that some non-trivial linear combination of rows is zero; the same linear combination of the b_i's is therefore also zero, #. ◆

Definition 4.9 The **dimension of V over F** is the size of any basis.

Corollary 4.10 (Basis Theorem—Finite Dimensional Version)
The following are equivalent for any collection of vectors $\mathcal{B}$ in the n-dimensional space V.
(a) $\mathcal{B}$ is a basis for V.
(b) $\mathcal{B}$ is a linearly independent spanning set of V.
(c) $\mathcal{B}$ is any collection of n linearly independent vectors in V.
(d) $\mathcal{B}$ is any collection of n vectors that generate V.

Proof See Exercise Set 4 #3 and 4, as well as the Basis Theorem. ◆

Exercise Set 4
1. Prove that, if $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ is a basis for V, and $v \in V$, then v can be expressed *uniquely* as a linear combination $c_1b_1 + c_2b_2 + \ldots + c_nb_n$ of the basis elements. The c_i are called the **coordinates of v with respect to $\mathcal{B}$**.
2. Prove that $F[x]$ is infinite dimensional over F.

3. Prove that, if V has a generating set consisting of n vectors, then any set of more than n vectors must be linearly dependent.
4. Find bases for all the vector spaces in Example 3.2 A-I. (Assume S is finite in example G.)
5. Let V be n-dimensional. Prove:
(a) C is any collection of n linearly independent vectors iff C is a maximal linearly independent set.
(a) C is any collection of n vectors that generate V iff C is a minimal generating set.

5. Linear Maps

Definition 5.1 Let V and W be vector spaces over F. A **linear transformation**, or **linear map from V to W** is a map f: $V \longrightarrow W$ such that $f(cv_1 + v_2) = cf(v_1) + f(v_2)$ for all $v_1, v_2 \in V$ and $c \in F$.

Note This is the same as requiring:
(a) $f(v_1 + v_2) = f(v_1) + f(v_2)$ for all $v_1, v_2 \in V$, and
(b) $f(cv) = cf(v)$ for all $v \in V$ and $c \in F$.

Examples 5.2
A. The zero map $V \longrightarrow W$ and the identity map on V are linear.
B. f: $\mathrm{R}^2 \longrightarrow \mathrm{R}^2$; $f(x, y) = (2x+3y, 4x-y)$.
C. f: $\mathrm{R} \longrightarrow \mathrm{R}$ is linear iff $f(x) = x.f(1) = mx$ for all x.
D. Let A be any $m \times n$ matrix, and define A_*: $F^n \longrightarrow F^m$ by $A_*(v) = A \cdot v$, where v is written as a column vector. This is the **linear map associated with a matrix.** We look at a formula for $f(v)$ when $v = (v_1, v_2, \ldots, v_n)$.
E. ε: $F[x] \longrightarrow F$; $\varepsilon(a_0 + a_1x + \ldots + a_nx^n) = a_0 + a_1x + \ldots + a_n$.
F. Differentiation on $C^\infty(\mathrm{R})$.

Definition 5.3 If f: $V \longrightarrow W$ is a linear map, then define its **kernel**, or **null space**, and **image**, or **range** as follows:
$\ker f = f^{-1}(0) = \{v \in V \mid f(v) = 0\}$
$\mathrm{im} f = f(V) = \{f(v) \mid v \in V\}$.
If f is injective, it is called a **monomorphism**. If it is surjective, it is called an **epimorphism**. If it is bijective, it is called an **isomorphism**.

Examples—look at all of the maps in Examples 5.2.

Lemma 5.4 (Kernel and Image are Subspaces)
If f: $V \longrightarrow W$ is a linear map, then $\ker f < V$ and $\mathrm{im} f < W$.

Exercise Set 5.

If V and W are finite dimensional, we can talk of the dimensions of $\ker f$ and $\mathrm{im} f$:

Definition 5.5 If f: $V \longrightarrow W$ is a linear map and V and W are finite dimensional, then we define the **nullity** of f to be dim(ker$f$), and the **rank** of f to be dim(Imf).

Theorem 5.6 (Rank Plus Nullity)
If f: $V \longrightarrow W$ is a linear map and V and W are finite dimensional, then
rank(f) + nullity(f) = dim(V).

Proof Let dim(V) = n, and nullity(f) = k, and let $\mathcal{B} = \{b_1, b_2, \ldots, b_k\}$ be a basis for ker$f$, so that $|B| = k$. By Theorem 4.8(b), $\mathcal{B}$ can be enlarged to a basis

$$\mathcal{B}' = \{b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_r\}$$

of V. Since it's enough to show that $C = \{f(c_1), f(c_2), \ldots, f(c_r)\}$ is a basis for Im(f), since $|C| = n-k$. But certainly, anything in the image of f is in the span of C (why?) so that C spans Im(f). Further, C is a linearly independent set, since if a linear combination of them is zero, the same linear combination of the c_i is in ker(f), which would contradict independence of the basis $\mathcal{B}'$ if the coefficients were not all zero. ◆

Definition 5.7 If A is an $m \times n$ matrix, then its **row rank** is the dimension of the subspace of F^n generated by the rows of A. Similarly, its **column rank** is the dimension of the subspace of F^m generated by the columns of A.

Note The row rank of a matrix A is just the rank of the associated linear map A_*. Thus, if we define the **nullity of the matrix A** to be the nullity of A_*, we have the following.

Corollary 5.8 (Rank Plus Nullity for Matrices)
If A is any $m \times n$ matrix, then
rank(f) + nullity(f) = n.

Corollary 5.9 (Row Rank = Column Rank)
If A is any $m \times n$ matrix, then its row rank equals its column rank.

Proof The column rank of A is just rank(A_*), the rank of the associated linear map A_*: $F^n \longrightarrow F^m$,. Also observe that its row rank is n − nullity(A_*) (see Exercise Set 5). That they are equal is thus simply a restatement of Corollary 5.8. ◆

Exercise Set 5

1. (Do not hand in) H&K, p. 73, #1-4.

2. (H&K, p. 73 #10) Give an example of a linear map C⟶C over R which is not linear over C.

3. Let f: $V \longrightarrow V$ be linear, and have the property that $f \circ f = f$. Show that ker(f)∩Im(f) = {0}, and that V = ker(f) + Im(f). (Such a linear map f is called a **projection.**)

4. Prove that a map f: $X \longrightarrow Y$ (between sets) is invertible iff it is bijective.

5. Prove that f: $V \longrightarrow W$ is a monomorphism iff ker(f) = {0}.

6. Prove Lemma 5.4.

7. Prove that if A is any $m \times n$ matrix, then its row rank = n − nullity(A_*).

6. Isomorphism

Recall that the linear map $f: V \longrightarrow W$ is a linear isomorphism if it is a bijective linear map. If there exists a linear isomorphism $f: V \longrightarrow W$, we write $f: V \cong W$, or simply $V \cong W$.

Note $f: V \longrightarrow W$ is an isomorphism iff $\ker(f) = \{0\}$ and $\mathrm{Im}(f) = W$. (See the last exercise set.)

Theorem 6.1 (Linearity of the Inverse)
If $f: V \longrightarrow W$ is a linear isomorphism, then its inverse is also linear (and hence a linear isomorphism as well.)

Proof in class.

Theorem 6.3 (Invertibility of Linear Map—Finite Dimensional Case)
If V and W are finite dimensional spaces with $\dim V = \dim W$, and if $f: V \longrightarrow W$ is linear, then the following are equivalent.
(a) f is invertible.
(b) f is a monomorphism.
(c) f is an epimorphism.

This is really a consequence of the rank plus nullity theorem (proof in class).

Theorem 6.4 (Classification of Finite Dimensional Vector Spaces Over F)
If V is any finite dimensional vector space over F, then $V \cong F^n$, where $n = \dim V$.

Proof in class.

Definition 6.5 Let $\mathrm{Hom}(V, W)$ be the set of linear maps $V \longrightarrow W$. Define an F-vector space structure on $\mathrm{Hom}(V, W)$ by taking the sum, $f+g$, of f, g in $\mathrm{Hom}(V, W)$ to be normal addition of maps, and $(cf)(x) = cf(x)$ for $c \in F$.

Recall that $M(m, n)$ is the vector space of $m \times n$ matrices over F.

Theorem 6.6 (Every Linear Map $F^n \longrightarrow F^m$ Comes From a Matrix)
Define $\phi: M(m, n) \longrightarrow \mathrm{Hom}(F^n, F^m)$ by $\phi(A) = A_*$. Then ϕ is a linear isomorphism. Further, ϕ and ϕ^{-1} both respect products: $\phi(AB) = \phi(A) \circ \phi(B)$ and $\phi^{-1}(f \circ g) = \phi^{-1}(f)\phi^{-1}(g)$. We call ϕ an **algebra isomorphism.**

Proof That ϕ is a linear map is routine to check. That it is an isomorphism follows from the fact that it has an inverse denoted by $f \mapsto [f]$, described in class. That it respects multiplication follows (essentially) from the fact that matrix multiplication is associative. ◆

Corollary 6.7
$f: F^n \longrightarrow F^m$ is invertible iff $[f]$ is invertible.

Corollary 6.8
If V and W are finite dimensional spaces, then $\mathrm{Hom}(V, W) \cong M(m, n)$, where $m = \dim W$ and $n = \dim V$.

Proof This follows from Theorems 6.4 and 6.6, and Exercise Set 6 #1.

Definition 6.9 If f: $V \longrightarrow W$ is linear, and $\mathcal{B}$ and $\mathcal{C}$ are bases of V and W respectively, then the **matrix of f with respect to the bases $\mathcal{B}$ and $\mathcal{C}$** is the matrix of f where we use the given bases to identify V with F^n and W with F^m. We write this matrix as $[f]_{\mathcal{CB}}$ (note the order). If $V = W$ and $\mathcal{B} = \mathcal{C}$, then we write $[f]_{\mathcal{BB}}$ as $[f]_{\mathcal{B}}$.

Note From the definition, it follows that the matrix of f is the matrix whose columns are the coordinates of the images of the elements of $\mathcal{B}$ with respect to the basis $\mathcal{C}$.

Examples 6.9 in class

Theorem 6.11 (Composition of Linear Maps)
If f: $V_{\mathcal{B}} \longrightarrow W_{\mathcal{C}}$, and g: $W_{\mathcal{C}} \longrightarrow U_{\mathcal{D}}$ are linear, then
$$[g \circ f]_{\mathcal{DA}} = [g]_{\mathcal{DC}}[f]_{\mathcal{CA}}.$$
(Here, $V_{\mathcal{X}}$ denotes V with the basis $\mathcal{X}$.)

Proof in the exercise set. (See Theorem 13 on p. 90 of H&K.)

Note Technically, this says that the associated matrix operation is a **functor** from the category of based vector spaces to the category of matrices.

Definition 6.12 If $\mathcal{B}$ and $\mathcal{C}$ are two bases for the f.d. v.s. V, then the **change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$** is defined to be $[1]_{BC}$; that is, the matrix of 1: $V_{\mathcal{C}} \longrightarrow V_{\mathcal{B}}$. (Note the reversal of convention.)

Note The columns of this matrix are the coordinates of the vectors in $\mathcal{C}$ with respect to the old basis $\mathcal{B}$.

Examples some in class.

Corollary 6.12 (Inverse of Change-of-Basis)
(a) If f: $V \longrightarrow W$ is an isomorphism, then its matrix (with respect to any choice of bases) is invertible.
(b) If $\mathcal{B}$ and $\mathcal{C}$ are any two bases for V, then $[1]_{CB} = [1]_{BC}^{-1}$.

Question Suppose f: $V_{\mathcal{B}} \longrightarrow W_{\mathcal{C}}$ has matrix A and $\mathcal{B}'$ and $\mathcal{C}'$ are new bases for V and W respectively. How is $[f]_{\mathcal{C'B'}}$ related to $[f]_{\mathcal{cb}}$?
Answer Consider the following commutative diagram of based vector spaces.

$$\begin{array}{ccc} V_{\mathcal{B}} & \xrightarrow{f} & V_C \\ \text{u}\,1 & & \text{u}1 \\ V_{\mathcal{B}'} & \xrightarrow{f} & V_{C'} \end{array}$$

This shows that

$$[f]_{C'\mathcal{B}'} = [1]_{cc'}^{-1}[f]_{C\mathcal{B}}[1]_{\mathcal{B}\mathcal{B}'}.$$

Notice that it says that, of P is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{B}'$ and if Q is the change-of-basis matrix from C to C', then the matrix of f with respect to the new bases is

$$Q^{-1}AP,$$

where A is the original matrix of f. In particular, if $\mathcal{B}$ and $\mathcal{B}'$ are two bases of V, and if f: $V \longrightarrow V$ is linear, then

$$\boxed{[f]_{\mathcal{B}'} = P^{-1}[f]_{\mathcal{B}}P,}$$

where P is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{B}'$.

Definition 6.13 The $n \times n$ matrices A and B are **similar** if there exists an invertible $n \times n$ matrix P with $B = P^{-1}AP$.

Exercise Set 6

1. Prove that, if $V \cong V'$ and $W \cong W'$, then $\text{Hom}(V, W) \cong \text{Hom}(V', W')$ as algebras over F.

2. Prove Theorem 6.11.

3. H&K p. 95 #1, 2, 5 (Do not hand in)

4. H&K, p. 95 #6 (Hand in)

5. Prove that, if V is any finite dimensional vector space over F, then $\text{Hom}(V, V)$ is also finite dimensional. What is its dimension?

Assignment Read up about the determinant.

7. Similar Matrices

We desire to put matrices in their simplest possible form by a change of basis. If A is a diagonal square matrix with diagonal entries d_i, then for every i, $A(e_i) = d_i e_i$. Thus we make the following definition.

Definition 7.1 If f: $V \longrightarrow V$ is any linear map on V, then an **eigenvector** is a nonzero vector v such that $f(v) = cv$ for some $c \in F$. The value $c \in F$ is called the corresponding **eigenvalue.** If A is any square matrix, then an **eigenvector** and corresponding **eigenvalue** of A are defined as one of A_*.

Proposition 7.1 (Finding Eigenvalues)
Let A be an $n \times n$ matrix. Then the following are equivalent.
(a) c is an eigenvalue of A.
(b) $A - cI$ is singular.
(c) $\det(A - cI) = 0$.

Proof c is an eigenvalue of A iff $(A-cI)v = 0$ for some non-zero v, and the latter is true iff $A-cI$ is singular, which is the same as saying $\det(A-cI) = 0$. ◆

Definition 7.2 If A is a square matrix, then the polynomial $\det(A-xI)$ is called the **characteristic polynomial** of A.

Note The determinant function is multiplicative, so that similar matrices have the same characteristic polynomial. (Why?)

Proposition 7.2 (Different Eigenvalues Give Independent Vectors)
If $\mathcal{V}$ is a collection of eigenvectors of A with distinct eigenvalues, then $\mathcal{V}$ is independent.

Proof By induction on the number of vectors in $\mathcal{V}$ (which we can assume finite). If one of them, v (with eigenvalue c) is a linear combination of the others, then we get

$0 = Av - cv =$ non-zero combination of fewer vectors,

a contradiction. ◆

Definition 7.3 The **eigenspace** of the eigenvalue $c \in F$ is the set $\{v \in V \mid Av = cv\}$.

Note that it is a subspace of V.

Proposition 7.4 (Dimension of Eigenspace)
If A is diagonalizable, then the dimension of the eigenspace associated with the eigenvalue c of A is the multiplicity of the root c in the characteristic polynomial of A.

Proof First note that the null space of $A - cI$ is the same as the null space of $A' - cI$ of A is similar to A'. Since A is diagonalizable, the null space of $A - cI$ is the same as the null space of $D - cI$, where D is diagonal. Thus its nullity k (which is also the dimension of the null space of c) is the number of occurrences of c on the diagonal (since the nullity of a diagonal matrix is the number of zeros on its diagonal). Since A and D have the same characteristic polynomial, and since the characteristic poly of D has $(x-c)$ repeating k times, we are done. ◆

Definition 7.5 The square matrix A s **diagonalizable** if it is similar to a diagonal matrix.

Note If A is diagonalizable, then the diagonalizing matrix P such that $P^{-1}AP$ is diagonal is the change-of-basis matrix from the standard basis to a basis of eigenvectors. This is the matrix whose columns are a basis of eigenvectors.

Example of diagonalizing a matrix in class: $A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$.

Theorem 7.6 (Criterion for Diagonalizability)
The following are equivalent for an $n\times n$ matrix A.
(a) A is diagonalizable.
(b) The dimensions of the eigenspaces add up to n.
(c) The characteristic polynomial of A has the form

$$(x-c_1)^{d_1}(x-c_2)^{d_2} \ \ldots\ (x-c_k)^{d_k}$$

where for each i, d_i is the dimension of the associated eigenspace.

Proof Exercise set 7 ◆

Exercise Set 7
1. Prove Theorem 7.6
2. Diagonalize the following matrices (from exercises in H&K)

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -10 & 8 & 7 \end{bmatrix}$$

3. Give an example of a non-diagonalizable matrix whose characteristic polynomial has the form given in Theorem 7.6(c). [Hint: Consider a simple upper triangular matrix.]
4. Give an example of a matrix that is diagonalizable over C but not over R.
5. Prove that, if A is similar to an upper triangular matrix, then the characteristic polynomial of A has the form

$$c(x) = (x-c_1)^{r_1}(x-c_2)^{r_2} \ \ldots\ (x-c_k)^{r_k},$$

with $\Sigma r_i = n$.

8. The Cayley-Hamilton Theorem

Quick Recollections from Ring Theory.
1. If R is a commutative ring (See H&K, p. 140), then an **ideal** in R is an additive subgroup J of R with the property that $rj \in J$ for every $r \in R$ and $j \in J$. The ideal J of R is **principal** if it has the form $\langle j\rangle = \{jr \mid r \in R\}$.
2. If $F[x]$ is the ring of polynomials over the field F, then $F[x]$ is a **principle ideal ring**; that is, every ideal J in $F[x]$ has the form $\langle p(x)\rangle$ for some polynomial $p(x) \in F[x]$. Further, we may take $p(x)$ to be the unique monic polynomial of minimal degree in J.
3. If V is any vector space, then Hom(V, V) is a ring.

Let T be any linear map on V, and let $p(x) \in F[x]$. Then we define an associated map $p(T)$: $V\longrightarrow V$ by taking

$$p(T) = a_0 + a_1T + \ \ldots\ + a_nT^n$$

for $\quad p(x) = a_0 + a_1x + \ \ldots\ + a_nx^n$.
(Here, a_0 is multiplication by a_0 as a map on T, and T^r is r-fold composition of T.)

Definition 8.1 The polynomial $p(x) \in F[x]$ **annihilates** the linear map T if the linear map $p(T) = 0$.

Examples 8.2

A. $T = 1$ iff $p(x) = x-1$ annihilates T.
B. Rotation in R^2 by a rational multiple of π is annihilated by x^n-1 for some n. So is any permutation of basis elements.
C. Every linear map T on a fd vector space V is annihilated by some polynomial. (**Proof**: the vector space Hom(V, V) is finite dimensional, whence some linear combination of the elements $1, T, T^2, \ldots, T^n$ must be independent…)
D. Let $V = F[x]$ and let T: $F[x] \longrightarrow F[x]$ be the shift operator (multiplication by x). Then T is annihilated by no polynomial in $F[x]$. (Why not? See the exercise set.)

Note The set of elements in $F[x]$ that annihilate T: $V \longrightarrow V$ is an ideal in $F[x]$. (Why?)

Definition 8.3 The **minimal polynomial** of T is the generator of the annihilator ideal of T. That is, it is the unique monic polynomial of minimal degree that annihilates T.

It follows that every polynomial that annihilates T is a multiple of the minimal polynomial.

Notes
1. We can think of matrices as linear maps, and therefore talk about the annihilating polynomial of a square matrix.
2. If $p(x)$ annihilates A, and B is similar to A, then $p(x)$ also annihilates B.

Theorem 8.4 (Minimal Polynomial of Diagonalizable Matrix)
If A is diagonalizable, then its minimal polynomial has the form $m(x) = (x-a_1)(x-a_2) \ldots (x-a_m)$, where a_i are the *distinct* eigenvalues of A.

(In particular, the minimal polynomial divides the characteristic polynomial).
Proof Let $p(x) = (x-a_1)(x-a_2) \ldots (x-a_m)$, where a_i are the distinct eigenvalues of T. If A is diagonalizable, then $p(x)$ certainly annihilates A, since any vector in R^n is a linear combination of eigenvectors, each of which is annihilated by $A-a_iI$ for some i. Further—and it suffices to show that—$p(x)$ is not divisible by any annihilating polynomial. Indeed, if $q(x)$ is a proper divisor of $p(x)$, then $q(x)$ is a product of some (but not all) of the factors $x-a_i$. If a_i is one of the missing ones, and if v is an associated eigenvector, then $q(A)(v) = \prod_{j\neq i}(a_i-a_j)v \neq 0$. ◆

More generally, we have:

Theorem 8.5 (Cayley-Hamilton)
If $c(x)$ is the characteristic polynomial of A, then $c(A) = 0$. In particular, $m(x) | c(x)$.

Proof Let K be the commutative ring subring of $M(n)$ consisting of all polynomials in A. (That is, K is the image of $F[x]$ under evaluation at A.) Let B be the $n\times n$ matrix with entries in K given by

$$B = \begin{bmatrix} A & 0 & \ldots & 0 \\ 0 & A & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & A \end{bmatrix} - \begin{bmatrix} a_{11}I & a_{21}I & \ldots & a_{n1}I \\ a_{12}I & a_{22}I & \ldots & a_{n2}I \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n}I & {}_{2n}I & \ldots & a_{nn}I \end{bmatrix}$$

$$= \begin{bmatrix} A-a_{11}I & -a_{21}I & \dots & -a_{n1}I \\ -a_{12}I & A-a_{22}I & \dots & -a_{n2}I \\ \vdots & \vdots & \vdots & \vdots \\ -a_{1n}I & -a_{2n}I & \dots & A-a_{nn}I \end{bmatrix},$$

where a_{ij} is the ij th entry of A. (Note the "negative" way it is written and the transpose of A.) Then $\det(B) = c(A)$, by definition of the characteristic polynomial $c(x)$ of A! (Note that we are replacing numbers by matrices when talking about $C(A)$.) Thus, we must prove that $\det(B) = 0$.

Now, we can compute directly that

$$B\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = 0,$$

the zero vector in $(F^n)^n$. On the other hand, if $\overline{B} = \text{adj}B$ is the formal adjoint of B (that is, the matrix whose entries are the transposes of the cofactors of the entries of B), so that

$$\overline{B}B = (\det B)I$$

(in K). Thus,

$$\det B\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \overline{B}B\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = 0.$$

But this can only happen if $\det B(e_i) = 0$ for each e_i. Now, $\det B$ is an $n \times n$ matrix with coefficients in the underlying field F, and the only way that $\det B(e_i)$ can be zero for each e_i is if $\det B$ is the zero matrix. Thus, $c(A) = \det(B) = 0$. ◆

Note Theorems 8.4 and 9.5, together with Exercise Set 8 #2 (below) tell us the following: Suppose that the characteristic polynomial $c(x)$ of A factors as

$$c(x) = (x-c_1)^{d_1}(x-c_2)^{d_2} \dots (x-c_k)^{d_k}.$$

Then the minimal polynomial has the form

$$m(x) = (x-c_1)^{r_1}(x-c_2)^{r_2} \dots (x-c_k)^{r_k},$$

where $1 \le r_j \le d_j$ for each j. In fact, we shall see that there is a matrix whose minimal polynomial is as above for any choices of r_j with $1 \le r_j \le d_j$ for each j.

If A happens to be diagonalizable, then, by 8.4, $m(x) = (x-c_1)(x-c_2) \dots (x-c_k)$, so that each $r_j = 1$. We shall see that the converse is also true: that is, if $m(x)$ has the above form, then A is diagonalizable.

In general, the goal is to say what the simplest form of A is, based on its minimal polynomial.

Exercise Set 8

1. Prove the assertion in Example 8.2(D).

2. Prove that $c(a) = 0$ iff $m(a) = 0$ for $a \in F$. (Here $c(x)$ denotes the characteristic polynomial of A, and $m(x)$ the minimal polynomial of A.) Deduce that $m(x)$ and $c(x)$ have the same distinct linear factors (although the powers may be different).
3. What does Hamilton-Cayley say about the relationship between $m(x)$ and $c(x)$ that is not already said in Exercise 2?
4. Find a 3×3 matrix whose minimal polynomial is x^3.
5. Prove that, if A is similar to an upper triangular matrix, then A has a characteristic polynomial of the form

$$m(x) = (x-c_1)^{r_1}(x-c_2)^{r_2} \ \ldots\ (x-c_k)^{r_k} .$$

9. Invariant Subspaces

Definition 9.1 The subspace W of V is invariant under the linear endomorphism f: $V\longrightarrow V$ (equivalently, it is A-invariant) if $f(W) \subset W$.

Examples 9.2

A. $\{0\}$, V are always invariant.
B. $f(V)$, $f(f(V)) = f^2(V)$, $f^3(V)$, . . . are always invariant (since $f(f(V)) \subset f(V)$), *etc.*)
C. $\ker f, \ker f^2, \ldots, \ker f^r$ are $f-$invariant.
D. The eigenspace associated with a particular eigenvector of A.
E. The subspace of V generated by all the eigenvectors (regardless of the eigenvalue).

Note If W is invariant under f, then there is a basis such that $[f]$ has block form

$$[f] = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} ,$$

where B is the matrix of $f|W$: $W\longrightarrow W$, the restriction of f to W.

Lemma 9.2 (Characteristic Polynomial of a Restriction)
If W is an f-invariant subspace, then the characteristic polynomial of $f|W$ divides the characteristic polynomial of f.

Proof Writing the matrix of f in the above block form, we see that

$$\det([f]-xI) = \det(B-xI)\det(A-xI),$$

proving the result. ◆

Definition 9.3 If f: $V\longrightarrow V$ is linear, $v \in V$, and $W < V$, then the ***f*−conductor of *v* into *W*** is

$$S_f(v;W) = \{p(x) \in K[x] \mid p(f)(v) \in W\}.$$

(We drop the subscript f when the map is understood.) When $W = 0$, we refer to $S(v;0)$ as the ***f*-annihilator of *v***.

Thus, thinking of f as a matrix A, $S(v;W)$ is the set of all polynomials $p(x)$ such that $p(A)$ sends v into W.

Examples 9.4

A. Let $v \in W$ and suppose W is A-invariant. Then $S(v;W) = K[x]$ itself.

B. If W is A-invariant and $v \in V$, then $S(v,W)$ is an ideal in $K[x]$ (since once you land in W, you stay there, by invariance.)

C. If v is an eigenvector of A with eigenvalue λ, then $x-\lambda \in S(v,0)$. Thus, since 0 is A-invariant, $S(v,0)$ is an ideal containing $x-\lambda$. Since that ideal must be generated by a unique monic polynomial of least degree in $S(v,0)$, and since $x-\lambda$ is a monic polynomial of least degree, it follows that $S(v,0) = \langle x-\lambda \rangle$. This is what the annihilators of eigenvectors look like.

D. Let W be A-invariant. Then, since the minimal polynomial $m(x)$ annihilates every $v \in V$, $m(x)$ must live in $S(v,0) \subset S(v,W)$. If $g(x)$ is the unique monic generator of $S(v,W)$, then $m(x)$ must be a (polynomial) multiple of it. In other words: *the generators of all conductors are all divisors of the minimal polynomial*.

Lemma 9.5 (Technical Lemma)
Suppose the minimal polynomial of A factors as

$$m(x) = (x-c_1)^{r_1}(x-c_2)^{r_2} \ \ldots\ (x-c_k)^{r_k},$$

and suppose that $W \subsetneqq V$ is A-invariant, then there exists an eigenvalue λ of A as well as a vector $v \in V-W$, with

$$(A-\lambda I)v \in W.$$

(For instance, v *could* be an eigenvector with eigenvalue λ, but not necessarily.) In other words, the A-conductor of v into W is generated by a monic polynomial of a particularly nice type.

Proof Pick any vector $u \in V-W$. Then, since W is A-invariant, we saw in Example 9.4(D) that its A-conductor is generated by a divisor of the minimal polynomial, so we can write this generator as

$$(x-c)s(x)$$

for some $s(x)$ and eigenvalue c (just look at $m(x)$ to see why.) Thus, let $v = s(A)(u)$. We need only establish that $s(A)(u) \notin W$. But if is was, then $s(x)$ would be a smaller degree element of the conductor than the generator. Done. ◆

What this lemma is good for is the following:

Theorem 9.6 (Characterization of Matrices Similar to Upper Triangular Matrices)
The following are equivalent for the $n \times n$ matrix A.
(a) A is similar to an upper triangular matrix.
(b) Its characteristic polynomial $c(x)$ is a product of linear factors.
(c) Its minimal polynomial $m(x)$ is a product of linear factors.

Proof
(a)⇒(b) This is Exercise Set 8 #5.

(b)⇒(c) This follows from the fact that $m(x)|c(x)$.

(c)⇒(a) Write $m(x) = (x-c_1)^{r_1}(x-c_2)^{r_2} \ . \ . \ . \ (x-c_k)^{r_k}$. We now construct the basis using the technical lemma. Start by applying it to $W = 0$. Thus, we get a v and an eigenvalue c with $(A-cI)(v) = 0$. This shows that $\{v\}$ is A-invariant, and the matrix of A with respect to any basis extending $\{v\}$ has the form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} ,$$

where B is diagonal (it happens to be 1×1 and consist of c at the moment). By induction, we assume we have constructed in A-invariant subspace W on dimension n such that A can be made to have the above form, with B upper triangular, and we must thus enlarge W. Applying the lemma one again gives an element v with $(A-cI)(v) \in W$, so that $Av \in \langle v\rangle + W$ This shows that $\langle v\rangle + W$ is A-invariant, so we add v to the present basis. Since now $A(v)$ has coordinate c in the $(n+1, n+1)$ spot, and other coordinates in $(j, n+1)$ spot with $j \le n$, we see that, with respect to the new basis, A still has the desired block form with B an $(n+1)\times(n+1)$ upper triangular matrix, thus completing the inductive step. ◆

Note By the proof, we can also say what the diagonal entries must be: the eigenvalues of A, of course!

Corollary 9.7 (Algebraically Closed Fields)
Every square matrix over an algebraically closed field (such as C) is similar to an upper triangular matrix.

Theorem 9.8 (Characterization of Diagonalizable Matrices)
The following are equivalent for the $n\times n$ matrix A.
(a) A is similar to a diagonal matrix.
(b) Its minimal polynomial $m(x)$ is a product of distinct linear factors $(x-c_i)$.
(c) Its characteristic polynomial $c(x)$ is a product of linear factors with the property that, if $(x-c_1), (x-c_2), \ . \ . \ . \ , (x-c_r)$ are the distinct factors, then

$$(A-c_1I)(A-c_2I). \ . \ . \ (A-c_rI) = 0.$$

Proof
(a)⇒(b) follows from Theorem 8.4.
(b)⇒(c) By Theorem 9.5, A must be triangulable, and so $c(x)$ must be a product of linear factors, showing the first part of the claim. Further, since $m(x) = (x-c_1)(x-c_2) \ . \ . \ . \ (x-c_r)$, where the c_i are the distinct eigenvalues (see Exercise Set 8 #2), it follows that

$$0 = m(A) = (A-c_1I)(A-c_2I). \ . \ . \ (A-c_rI).$$

(c)⇒(a) The hypothesis implies that the minimal polynomial $m(x)$ must be $(x-c_1)(x-c_2) \ . \ . \ . \ (x-c_r)$, since it must be at least as big as that, and it does annihilate A.

Now, if A was not diagonalizable, then the subspace W generated by all the eigenvalues (which we saw was invariant) would not equal the whole of V. By the technical lemma, there is an eigenvalue c and a vector v not in W with $Av - cv \in W$, and is thus a linear combination of eigenvectors. In other words,

$$(A-cI)v = v_1 + \ . \ . \ . \ + v_n \qquad (1)$$

for some eigenvectors v_i. Now let $q(x)$ be $m(x)$ with the factor $(x-c)$ missing. Then, since $m(A) = 0$,we get

$$0 = m(A)(v) = (A-cI)q(A)v.$$

Thus, since $q(A)v$ is killed by $A-cI$, it must be an eigenvector, and hence in W. (2)
Now the polynomial $q(x) - q(c)$ is divisble by $x-c$, so write

$$q(x) - q(c) = h(x)(x-c) \qquad (3)$$

Also,

$$\begin{aligned} q(A)v &= (q(A)-q(c)I)v + q(c)v \\ &= h(A)\cdot(A-cI)v + q(c)v && \text{by (3)} \\ &= h(A)((v_1 + \ldots + v_n) + q(c)v && \text{by (1)} \end{aligned}$$

But the left-hand side is in W by (2), and so is the firs summand on the right. Thus, $q(c)v \in W$. Thus either the scalar $q(c) = 0$ (implying that $(x-c)$ is a root of $q(x)$, and hence a double root of $m(x)$—a contradiction), or else v itself must be in W—another contradiction. ◆

Exercise Set 9

1. Prove that any matrix is similar to a matrix of the form

$$\begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where B is diagonal $r\times r$ matrix, where r is the sum of the dimensions of all the eigenspaces. (Hint: See Example 9.2 E above.)

2. p. 205 # 4.

4. (Based on p. 205 # 7 &5)

(a) Show that if A is diagonalizable iff it annihilated by some polynomial that factors into linear terms with distinct roots.

(b) Deduce that every matrix A such that $A^2 = A$ is similar to a diagonal matrix.

5.

(a) Show that if A is any 2×2 non-diagonal matrix with real entries, then A is diagonalizable iff $(\text{tr}A)^2 > 4\det A$.

(b) Show that if A is any 2×2 non-diagonal matrix with complex entries, then A is diagonalizable iff $A\ (\text{tr}A)^2 \neq 4\det A$.

6. (p. 206 #11) Prove or disprove: If an upper triangular matrix is similar to a diagonal matrix, then it is already diagonal.

10. Inner Product Spaces

Note From now on, F = R or C.

Definition 10.1 An **inner product** on the vector space V over F is a map

$$(-|-)\colon V\times V \longrightarrow F$$

such that, for all vectors $v, w, u \in V$ and $c \in F$, one has:

$$\left.\begin{aligned} (v+w|u) &= (v|u) + (w|u) \\ (v|w+u) &= (v|w) + (v|u) \\ (cv|w) &= c(v|w) \\ (v|cu) &= \bar{c}(v|u) \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(bilinearity)}$$

$(v|w) = \overline{(w|v)}$ .....(symmetry)

$(v|v)$ is real and ≥ 0 and $(v|v) = 0$ iff $v = 0$ ...(positive definite)

An **inner product space** is a vector space with an inner product. If $F = \mathrm{R}$, we refer to a **Euclidean space**, and if $F = \mathrm{C}$, to a **unitary space.**

Note The reason for the complex conjugagtes is that we want $(v|v)$ to come out to be real. (Look at what would happen to the porperties if we changed the definition of the standard inner product on C^n below.)

Examples 10.2

A. The standard inner product on R^n given by $(v|w) = \Sigma v_i w_i$

B. The standard inner product on C^n given by $(v|w) = \Sigma v_i \bar{w}_i$.

C. Define an inner product on $M(n;\mathrm{C})$ by $(A|B) = \mathrm{tr}(A\bar{B})$, where tr is trace. (Note that this concides with the "standard" one if we note that $(A|B) = \Sigma_{i,j} A_{ij}\bar{B}_{ij}$.)

D. If $V = C[0, 1]$, the vector space of all continuous complex-valued functions on $[0, 1]$, then define

$$(f|g) = \int_0^1 f(x)\overline{g(x)}\, dx \ .$$

E. New inner products from old: If $(-|-)$ is an inner product on V and if f: $V \longrightarrow V$ is linear, then we ge a new inner product, $(-|-)_f$ given by

$(v|w)_f = (f(v)|f(w))$.

F. Every unitary space inherits the structure of a Euclidean space.

Definition 10.3 If $(V, (-|-))$ is an inner product space and $v \in V$, define the **norm** of v to be

$\|v\| = (v|v)^{1/2}$.

(In the one-dimensional case, this is the usual length function.)

Note $\|v \pm w\|^2 = \|v\|^2 \pm 2\mathcal{R}e(v|w) + \|w\|^2$ for all $v, w \in V$.

Proposition 10.4 (Properties of the Norm)

Let $(V, (-|-))$ be an inner product space and let $v, w, u \in V$ and $c \in F$. Then

(a) $\|v\| \geq 0$, and equals 0 iff $v = 0$. (positive definite)

(b) $\|cv\| = |c| \cdot \|v\|$, where $|-|$ denotes (standard) magnitude.

(c) $|(v|w)| \leq \|v\| \cdot \|w\|$

(d) $\|v+w\| \leq \|v\| + \|w\|$ (Cauchy-Schwartz)

Proof

(a) This is an immediate consequance of the positive definite property for inner products.

(b) Immediate from definition and bilinearity ($c\bar{c} = |c|^2$).

(c) Use the fact that $(x|x) \geq 0$, where $x = v - \dfrac{(v|w)}{\|w\|^2}\ w$ (*if* $w \neq 0$).

(d) now follows from (c) by squaring both sides, since $\mathcal{R}e(v|w) \leq |(v|w)|$. ◆

Definition10.5 The vectors v and w are **orthogonal** if $(v|w) = 0$. A set of vectors that are mutually orthogaonal is an **orthogonal set**. An **orthonormal set** of vectors is an orthogonal set each of whose vectors has norm 1. Thus, if v and w are in an orthonormal set, then

$$(v|w) = \begin{cases} 1 & \text{if } v = w; \\ 0 & \text{if } v \neq w. \end{cases}$$

Examples 10.6

A. Standard o/n basis on R^n or C^n.

B. Fourier Series: If $V = C([0,\ 1],\ \mathrm{R})$ the vector space of all real-valued functions on $[0, 1]$, with

$$(f|g) = \int_0^1 f(x)g(x)\ dx\ ,$$

then $\{1,\ \sqrt{2}\ \sin 2\pi nx\ \ \sqrt{2}\ \cos 2\pi nx \mid n \geq 1\}$ is an orthonormal set.

C. Complex Fourier Series: If $V = C[0,\ 1]$, the vector space of all continuous complex-valued functions on $[0, 1]$, with

$$(f|g) = \int_0^1 f(x)\overline{g(x)}\ dx\ ,$$

then $\left\{\frac{1}{\sqrt{2}}\, e^{2\pi inx} \;\middle|\; n \geq 0\right\}$ is an orthonormal set too. (This is actually obtained by taking the linear combinations such as $\frac{1}{\sqrt{2}}\left(\sqrt{2}\cos 2\pi nx + i\sqrt{2}\sin 2\pi nx\right)$ of the vectors in **B**.

D. Gram-Schmidt: Every inner product space has an orthonormal basis. (We do the finite dimensional case, but induction—or transfinite induction—also workds for infinite dimensional cases.)

Proposition 10.7
Orthogonal sets are linearly independent.

Definitions 10.8 (a little different from the text) If U is an inner product space and $V < U$ and $u \in U$, then define

$\pi_V\colon U \longrightarrow V$ by $\pi_V(u) = \Sigma_i (u|v_i)v_i \in V$.

where $\{v_i\}$ is any orthonrmal basis of V. (Note that the sum is always finite.) The map π_V is called **orthogonal projection onto V.**

Lemma 10.8
This definition is independent of the choice of orthonormal basis for V.

Proof Let $\{v_i\}$ and $\{w_j\}$ be any two o/n bases for V, and denote $(x|v_i)|v_i$ (summation convention in force) by $\pi_V(x)_1$ and $(x|w_i)|w_i$ by $\pi_V(x)_2$.

Claim 1: For all j, one has $\pi_V(w_J)_1 = \pi_V(w_J)_2$

Indeed, write $w_j = w_{jk}v_k$ (using basis to expand w_j with $w_{jk} \in F$. Then

$$\pi_V(w_J)_1 = (w_j|v_i)v_i$$

$$= (w_{jk}v_k|v_i)v_i$$
$$= w_{jk}(v_k|v_i)v_i$$
$$= w_{jk}\delta_{ki}v_i$$
$$= w_{ji}v_k$$
$$= w_j = \pi_V(w_J)_2.$$

Claim 2: For all $x \in U$, one has $\pi_V(x)_1 = \pi_V(x)_2$.

This is in the exercises.

◆

Definition 10.9 Let $V < U$, where U is an inner product space. The **orthogonal complement** of V in U is

$$U-V = V^{\perp} = \{u \in U \mid (u|v) = 0 \text{ for all } v \in V\}.$$

If V has orthonormal basis $\{v_i\}$, then the **orthogonal projection** from U onto V is given by

$$\pi_V\colon U \longrightarrow V;\ \pi_V(u) = u - \Sigma_i(u|v_i)|v_i$$

Proposition 10.10 (Orthogonal Decomposition)
If U is any inner product space and $V < U$, then orthogonal projection decomposes U as a direct sum,

$$U = V \oplus V^{\perp}$$

Proof By construction, π_V has the property that $\pi_V{}^2 = \pi_V$. ◆

Exercise Set 10

1. Verify that the inner product in Example 10.2(E) is indeed an inner product.

2. Show that every inner product on V is entirely determined by its "real part," $\mathcal{R}e\circ(-|-)$, where $\mathcal{R}e\colon C \longrightarrow R$ is given by taking the real part.

3. Show that, for all $v, w \in V$,

$$(v|w) = \frac{\|v+w\|^2 - \|v-w\|^2 + i[\|v+iw\|^2 - \|v-iw\|^2]}{4}\ .$$

4. If is a complex inner product space, we can also regard it as a real inner product space by composing the inner product with $\mathcal{R}e\colon C \longrightarrow R$. Show that this defines a one-to-one correspondence between unitary structures and Euclidean structures on V.

5. Parallelogram Law Show that, for all $v, w \in V$,

$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2.$$

6. Prove Claim 2 in Lemma 10.8. [Expand the basis $\{w_j\}$ to an orthonormal basis $\{w_j, \mu_j\}$ for U…]

11. Adjoint Operators

Definition 11.1 A **linear functional** on the vector space V over F is a linear map $f\colon V \longrightarrow F$.

Examples 11.2

A. Let $f\colon R^n \longrightarrow R$ be given by $f(x_1, x_2, \ldots, x_n) = \Sigma_i a_i x_i$, where $a_i \in F$ are arbitary.

B. If V is an inner product space and $u \in V$, define T_u: $V \longrightarrow F$ by $T_u(v) = (v|u)$. (Why didn't we put v in the second slot?)

Note We are not giving any more examples since, as we shall see, every example is of type B. In fact:

Theorem 11.3 (Representability of Linear Functionals)
Let V be a finite dimensional inner product space over F, and let f be a linear funtional. Then there exists $u \in V$ such that $f = T_u$. Further, u is the unique vector with this property.

Proof Choose an orthonormal basis $\{e_1, e_2, \ldots, e_n\}$ of V, so that

$$f(v) = f(\Sigma_i v_i e_i) = \Sigma_i v_i f(e_i) = \Sigma_{i,j} v_i \delta_{i,j} f(e_j) = \Sigma_{i,j} v_i (e_i|e_j) f(e_j)$$
$$= \Sigma_j (v|e_j) f(e_j) = (v|\Sigma_j \overline{f(e_j)} e_j) = (v|u),$$

where

$$u = \overline{f(e_j)} e_j.$$

Uniqueness is left to the exercises. ◆

Theorem 11.4 (Existence of the Adjoint)
Let V be a finite dimensional inner product space over F, and let f: $V \longrightarrow V$ be linear. Then there exists a unique linear map f^*: $V \longrightarrow V$ such that, for all $u, v \in V$,

$$(f(u)|v) = (u|f^*(v)).$$

Further, one has $f^{**} = (f^*)^* = f$, so that f is the adjoint of f^*.

Proof For each $v \in V$, note that the assignment $u \mapsto f(u) \mapsto (f(u)|v)$ is a linear functional, whence, by Theorem 11.3, there exists a unique $w_v \in V$ such that $(f(u)|v) = (u|w_v)$. Define $f^*(v) = w_v$. We check that f^* is linear in class.

For the second part, one has

$$(f^*(u)|v) = \overline{(v|f^*(u))} = \overline{(f(v)|u)} = (u|f(v)),$$

showing that $f^{**} = f$. ◆

Proposition 11.6 (Matrix of a Linear Operator And Its Adjoint)
If f: $V \longrightarrow V$ is a linear operator on the inner product space V, then, if $\mathcal{B} = \{e_i\}$ is an orthonormal basis, one has

$$[f]_{\mathcal{B}} = [(f(e_i)|e_j)]$$

and

$$[f^*]_{\mathcal{B}} = [\ \overline{(f(e_j)|e_i)}\ ] = \overline{[f]_{\mathcal{B}}^T}\ .$$

Proof The columns of $[f]_{\mathcal{B}}$ have, as entries, the coordinates of $f(e_i)$ with respect to $\mathcal{B}$. But, with $f(e_i) = \Sigma_j a_{ji} e_j$, taking $(-|e_j)$ of both sides of the equation gives $a_{ji} = (f(e_i)|e_j)$, showing the first part. For the second part, the ji entry of f^* is given by

$$(f^*(e_i)|e_j) = (e_i|f(e_j))$$
$$= \overline{(f(e_j)|e_i)}\ ,$$

as required. ◆

Definition 11.7 The $n\times n$ matrix A is **self-adjoint** or **Hermitian** if $A = \overline{A}^T$.

If follows that every self-adjoint matrix is the matrix of a self-adjoint linear operator.

Exercise Set 11
1. Complete the proof of Theorem 11.3.
2. Prove Theorem 9 on p. 207 without looking.

12. Unitary Operators

Definition 12.1 If V and W are inner product spaces, then a **linear isometry** f: $V\longrightarrow W$ is a linear map that preserves the inner products. That is, $(f(u) \mid f(v)) = (u \mid v)$ for all $u, v \in V$.

Remarks
1. Linear isometries preserve the norm: $\|f(v)\| = \|v\|$ for all $v \in V$.
2. By (1), it follows that linear isometries are injective. Thus, if V and W are finite dimensional, then a linear isometry $V\longrightarrow W$ is a linear isomorphism.

The following result states that it is sufficient to look at what happens to a basis, and gives us many examples.

Proposition 12.2 (Criterion for a Linear Isometry)
Let V and W be finite dimensional spaces over R or C. Then the following conditions on f: $V\longrightarrow W$ are equivalent.
(a) f is a linear isometry.
(b) The image of every orthonormal basis under f is an orthonormal basis.
(c) The image of some orthonormal basis under f is an orthonormal basis.
(d) f preserves the norm.

Proof
(a)⇒(b) *a fortiori*
(b)⇒(c) *a fortiori*
(c)⇒(d) Let $\{e_i\}$ be a basis whose inage under f is orthonormal. Then, if $u \in V$, we have(using summation convention)

$$\|f(u)\|^2 = (f(u) \mid f(u)) = (f(u_ie_i) \mid f(u_je_j)) = u_i\overline{u}_j(f(e_i) \mid f(e_j))$$
$$= u_i\overline{u}_j(e_i \mid e_j) = (u_ie_i \mid u_je_j) = (u \mid u) = \|u\|^2.$$

(c)⇒(d) Let $u, v \in V$. By Exercise Set 10 #3, we can express $(u \mid v)$ as a linear combination of $\|u\pm v\|^2$ and $\|u\pm iv\|^2$. The result now follows. ◆

Corollary 12.3
(a) V and W are linearly isometric iff they have the same dimension.
(b) If V is given any orthonormal basis, then the matrix A represents a linear isometry iff its columns form an orthonormal basis.
(c) If V is given any orthonormal basis, then the matrix A represents a linear isometry iff its inverse is its adjoint.

(a) follows since one can use orthonormal bases to construct an inverse.

(b) follows from condition (c) in the proposition.
(c) follows from (b) using the definition of the inner product and matrix multiplication. ◆

Definition 12.4 A linear isometry f: $V \longrightarrow V$ is called a **unitary operator**, and the matrix of a unitary operator with respect to an orthonomal basis is called a **unitary matrix**. In the real case, we call a unitary matrix an **orthogonal matrix.**

Notes
1. It follows that A is a unitary matrix iff $A^{-1} = \overline{A}^T$, or, equivalently, $A\overline{A}^T = I$, or, equivalently, its columns are an orthonormal basis. In the real case, this is the same as saying that its transpose is its inverse.
2. Further, if f is unitary, its inverse is its adjoint (since they have the same matrix with respect to any orthonormal basis) so that:

$$((f(u) \mid v) = (u \mid f^*(v)) = (u \mid f^{-1}(v)).$$

3. Note that any unitary matrix commutes with its adjoint (since the adjoint is the inverse).

Definition 12.5 A square matrix that commutes with its adjoint is called **normal**.

Examples All unitary and self-adjoint matrices.

Theorem 12.5 (Diagonalizability of Complex Normal Matrices)
If A is a normal matrix, then there is a unitary matrix P such that $P^{-1}AP$ is diagonal.

Proof
A Special Case: A is self-adjoint (Hermitian) We do induction on dim$V$. If dim$V = 1$, then the result is obvious, since we are talking 1×1 matrices. If dim$V = n+1$, then any n-dimensional $A-$invariant subspace will have a basis of eigenvectors, by induction. If $F =$ C, then we can get such an invariant space as follows: choose an eigenvector v (there is always one, since we are working over C) of unit length, and then let W be the orthogonal complement of $\langle v\rangle$. Since A is self-adjoint, $(Aw \mid v) = (A^*w \mid v) = (w \mid Av) = (w \mid \lambda v) = 0$ for all $w \in W$, showing that W is A-invariant, as claimed. Hence we are done, since we can now choose an orthonormal basis of eigenvectors for W by induction. (Note that P, being the matrix whose columns form this orthonormal basis, is automatically unitary.)
The General Case (which does not need the Special Case) First, we claim that A can be made upper triangular using a unitary matrix. That is, there is an orthonormal basis with respect to which A is upper triangular. We again do induction on the dimension, starting with the easy one-dimensional case, choose an eigenvector v of A^* this time (we can, since we are working over C), and let W be its orthogonal complement. Then W is seen to be A-invariant (although $\langle v\rangle$ may not be). Choosing a basis that makes $A|W$ upper triangular, we then add the vector v as the last vector, getting the claim.
Claim We now claim: If A an upper triangular matrix, and A is normal, then A is diagonal. But, if A commutes with its adjoint, then we have

$A^*A = AA^*$, so that
$A\overline{A}^T = \overline{A}^TA$.

Equating the 1,1 entries and the n,n entries of $A\bar{A}^T$ and $\bar{A}^T A$ shows that A must have the form

$$A = \begin{bmatrix} a & 0 & 0 & \dots & 0 \\ 0 & b_1 & \dots & * & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{n-2} & 0 \\ 0 & 0 & \dots & 0 & c \end{bmatrix},$$

since the off diagonal terms give sums of the form $x\bar{x} = \|x\|^2$, and hence must each be zero. This gives the result by induction. ◆