LINEAR ALGEBRA 2

DAVID TWEEDLE

These notes will serve as reference for Math 3273. The notes follow [Axl15] and [GH17]. Thank you for reading.

1. Row, column, nullspace, left-nullspace and rank-nullity theorem

Let A be an n-by-m matrix with entries in a field k. You can think of k being either $\mathbb{R}$ or $\mathbb{C}$. A row vector is $(x_1, x_2, \ldots, x_m)$. A column vector is

$$\mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and the transpose of $\mathbf{v}$ is the row vector $\mathbf{v}^T = (x_1, x_2, \ldots, x_n)$. The set of all column vectors consisting of n entries in the field k will be denoted k^n.

Definition 1.1. The nullspace of A, denoted $\mathrm{null}(A)$, is the set of all column vectors $\mathbf{v} \in k^n$ such that $A\mathbf{v} = \mathbf{0}$.

Definition 1.2. The rowspace of A, denoted $\mathrm{row}(A)$, is defined to be the set of all linear combinations of the rows of A.

Definition 1.3. The columnspace of A, denoted $\mathrm{col}(A)$, is defined to be the set of all linear combinations of the columns of A.

Definition 1.4. The left-nullspace of A, denoted $\mathrm{null}(A^T)$, is defined to be the set of all vectors $\mathbf{y}$ such that $\mathbf{y}^T A = \mathbf{0}^T$.

Proposition 1.5. *Suppose A and B are both m-by-n matrices and ϵ is an elementary row operation. Suppose also that $A \xrightarrow{\epsilon} B$ and $I \xrightarrow{\epsilon} E$. Then $B = EA$.*

Theorem 1.6. *Let A be an m-by-n matrix with entries in the field k. Then there exists an invertible matrix P and a reduced row-echelon matrix R such that $A = PR$.*

Date: September 5, 2021.

Definition 1.7. The nullity of a matrix A, denoted $\text{nullity}(A)$, is defined to be the dimension of the nullspace of A.

The rank of A, denoted $\text{rank}(A)$, is defined to be the dimension of the row space of A.

Theorem 1.8. *Let A be an m-by-n matrix. Then*

$$\text{rank}(A) + \text{nullity}(A) = n.$$

2. Direct sums, basis, dimension

Definition 2.1. Let U and W be subspaces of a vector space V. Suppose

a) $U + W = V$
b) $U \cap W = 0$.

Then we say that V is the direct sum of U and W and write $V = U \oplus W$.

Definition 2.2. Let $S \subseteq V$. A linear combination of vectors from S is a vector $\mathbf{v}$ such that

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{v}_i$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ and $c_1, c_2, \ldots, c_n \in k$.

The span of S, denoted $\text{span}(S)$, is defined to be the set of all linear combinations of vectors from S.

Definition 2.3. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a list of vectors from V. Consider the equation

$$\sum_{i=1}^{n} c_i \mathbf{v}_i = \mathbf{0}.$$

The trivial solution is $c_1 = c_2 = \cdots = c_n = 0$.

The list of vectors is called linearly independent if the only solution to the equation is the trivial solution. Otherwise, the list is called linearly dependent.

Definition 2.4. Let S be a set of vectors. Then S is called linearly dependent if for any positive integer $n \geq 1$, any list of n distinct vectors from S is linearly independent.

Otherwise, S is called linearly dependent.

Definition 2.5. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a list of linearly independent and spanning vectors for V then we say that the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis for V.

The notion of an infinite basis is as follows: S is a basis for V if $\text{span}(S) = V$ and S is linearly independent.

Theorem 2.6. *Suppose* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is a basis for* V. *If* $\sum_{i=1}^n c_i\mathbf{v}_i = \sum_{i=1}^n d_i\mathbf{v}_i$ *then* $c_i = d_i$ *for all* $i = 1, 2, \ldots, n$.

Proposition 2.7. *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *be a list of vectors in* V. *Suppose* $\mathbf{v} = \sum_{i=1}^n c_i\mathbf{v}_i$ *with* $c_j \neq 0$ *for some* $1 \leq j \leq n$.

a) *Replacing* $\mathbf{v}_j$ *with* $\mathbf{v}$ *does not change the span of the list.*
b) *If the list* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is linearly independent, then it is still linearly independent after replacing* $\mathbf{v}_j$ *with* $\mathbf{v}$
c) *If* $\mathbf{v} \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_u)$ *for some* $u < n$, *then there exists* $j > u$ *with* $c_j \neq 0$
d) *If* $\mathbf{v} \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *then* $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}$ *is linearly independent.*

Theorem 2.8. *Let* $W \subset V$ *be a subspace of* V. *Suppose* V *has a finite basis* $\mathbf{v}_1, \ldots, \mathbf{v}_n$. *Suppose* $\mathbf{w}_1, \ldots, \mathbf{w}_k \in W$ *are linearly independent. Then there exists* $\mathbf{w}_{k+1}, \ldots, \mathbf{w}_u \in W$ *such that* $\mathbf{w}_1, \ldots, \mathbf{w}_u$ *is a basis for* W *and* $u \leq n$.

Definition 2.9. Let V be a vector space with basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then the dimension of V, denoted $\dim(V)$, is defined to be n.

If $V = 0$ then we write $\dim(V) = 0$ (or check that the empty list is linearly independent by definition, and spanning by convention).

If V has no finite basis (equivalently, V contains an infinite linearly independent set), then write $\dim(V) = \infty$.

Theorem 2.10. *Let* V *be a vector space and* $U \subset V$ *a subspace. Then there is a subspace* W *such that* $V = W \oplus U$.

Theorem 2.11. *Let* V *be a vector space with subspaces* $V_1, \ldots, V_n$. *Then the following are equivalent*

a) $V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$
b) *every* $\mathbf{v} \in V$ *can be written uniquely as* $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_n$ *where* $\mathbf{v}_i \in V_i$

Proof. Omitted. □

3. Linear transformations and matrices

Definition 3.1. Let V be a vector space with basis $B = \mathbf{v}_1, \ldots, \mathbf{v}_n$. Let $\mathbf{v} \in V$. Then there are unique scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{v} = \sum_{i=1}^n c_i\mathbf{v}_i.$$

We define the coordinate vector of $\mathbf{v}$ relative to B, denoted by $[\mathbf{v}]_B$, to be $[\mathbf{v}]_B = (c_1, c_2, \ldots, c_n)^T$.

Proposition 3.2. *Let V be a vector space and suppose that V can be written as the direct sum of subspaces*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_n.$$

Suppose that $B_1 = \mathbf{v}_1, \ldots, \mathbf{v}_{r_1}$ is a basis for V_1, $B_2 = \mathbf{v}_{r_1+1}, \ldots, \mathbf{v}_{r_1+r_2}$ is a basis forV_2, ..., $B_n = \mathbf{v}_{r_1+\cdots+r_{n-1}+1}, \ldots, \mathbf{v}_{r_1+\cdots+r_n}$ is a basis for V_n. Then $B_1 \cup B_2 \cup \cdots \cup B_n$ is a basis for V, and for all $\mathbf{v} \in V$ the coordinate vector of $\mathbf{v}$ may be written as

$$\left(\begin{array}{c} \mathbf{x}_1 \\ \hline \mathbf{x}_2 \\ \hline \vdots \\ \hline \mathbf{x}_n \end{array}\right)$$

where $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_n$ is the decomposition from Theorem 2.11 where $\mathbf{v}_i \in V_i$ and $\mathbf{x}_i = [\mathbf{v}_i]_{B_i}$.‘

Proof. Omitted. □

Proposition 3.3. *Let $B = \mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of a vector space V. Then*

a) for all $\mathbf{v}, \mathbf{w} \in V$,

$$[\mathbf{v} + \mathbf{w}]_B = [\mathbf{v}]_B + [\mathbf{w}]_B$$

b) for all $\mathbf{v} \in V$ and $c \in k$

$$[c\mathbf{v}]_B = c[\mathbf{v}]_B.$$

Proof. Omitted. □

Definition 3.4. Let V and W be vector spaces (always over a common field k). Let $T : V \to W$ be a function such that

a) $T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}')$ for all $\mathbf{v}, \mathbf{v}' \in V$

b) $T(c\mathbf{v}) = cT(\mathbf{v})$ for all $c \in k$, $\mathbf{v} \in V$.

Then T is called a linear transformation from V to W. If $V = W$ then $T : V \to V$ is called a linear operator on V.

Definition 3.5. Let k be a field. An m-by-n matrix over k is an array of mn elements of k arranged into m rows and n columns:

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

The set of all m-by-n matrices over k is denoted by $k^{m \times n}$.

Definition 3.6. Let $T : V \to W$ be a linear transformation. Let $B = \mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis for V and $B' = \mathbf{w}_1, \dots, \mathbf{w}_m$ be a basis for W. For each $j = 1, 2, \dots, n$ write

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i$$

and then define the matrix of T relative to the bases B and B', denoted ${}_{B'}[T]_B$, by

$${}_{B'}[T]_B = [a_{ij}] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Proposition 3.7. *Let $T : V \to W$ be a linear transformation. Let $B = \mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis for V and $C = \mathbf{w}_1, \dots, \mathbf{w}_m$.*

Then for all $\mathbf{v} \in V$

$$[T(\mathbf{v})]_C = {}_C[T]_B[\mathbf{v}]_B$$

Proof. Omitted. □

Definition 3.8. If A is an m-by-n matrix over a field k, define a linear transformation $T_A : k^n \to k^m$ by the rule

$$T_A(\mathbf{v}) = A\mathbf{v}$$

for all $\mathbf{v} \in k^n$.

Definition 3.9. Let V be a vector space. Define the identity function, denoted I_V (or just I if V is understood), by the rule

$$I_V(\mathbf{v}) = \mathbf{v} \text{ for all } \mathbf{v} \in V.$$

Clearly, I_V is a linear transformation.

Definition 3.10. Let $n \geq 1$ be an integer and let k be a field.

For each $i = 1, 2, \dots, n$ define the vector $\mathbf{e}_i \in k^n$ to be a column vector with a 1 in the i-th row and 0 everywhere else.

Then $B = \mathbf{e}_1, \dots, \mathbf{e}_n$ is called the standard basis of k^n.

Define the identity matrix to be the n-by-n matrix I such that

$$I = [\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n]$$

In other words, I is the n-by-n matrix with 1's along the diagonal and 0's everywhere else.

4. Invertibility of matrices, and determinants

Definition 4.1. Let $T : V \to W$ and $S : U \to V$ be two linear transformations. Define $T \circ S : U \to W$ by the rule

$$(T \circ S)(\mathbf{u}) = T(S(\mathbf{u})) \text{ for all } \mathbf{u} \in U.$$

Check that $T \circ S$ is a linear transformation.

Then check that the function

$$\Phi : \mathcal{L}(U, V) \to \mathcal{L}(U, W)$$

defined by $\Phi(S) = TS$ is a linear transformation.

Similarly for $\Psi(T) = TS$ where

$$\Psi : \mathcal{L}(V, W) \to \mathcal{L}(U, W).$$

Definition 4.2. Let $T : V \to W$ be a linear transformation. If there is a linear transformation $S : W \to V$ such that $ST = I_V$ and $TS = I_W$, then T is called invertible.

If T is invertible, then the inverse is unique, and we write $S = T^{-1}$.

Furthermore, T is invertible if and only if T is 1-1 and onto.

Definition 4.3. Let $n \geq 1$. Then there is a (unique) function $\det : k^{n\times n} \to k$ such that

a) $\det(\mathbf{v}_1, \ldots, \mathbf{v}_j + \alpha\mathbf{v}, \cdots, \mathbf{v}_n) =$
$\det(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) + \alpha \det(\mathbf{v}_1, \ldots, \mathbf{v}, \ldots, \mathbf{v}_n)$

b) $\det(\ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots) = -\det(\ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots)$

c) $\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$

Proposition 4.4. *Let A be an n-by-n matrix. Let A_{ij} be the matrix obtained by deleting the i-th row and j-th column of A. Then for any $1 \leq j \leq n$*

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

and for any $1 \leq i \leq n$

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

Proposition 4.5. *Let A be an n-by-n matrix. Then*

$$|A| = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

Theorem 4.6. *Let A be an n-by-n matrix. The following are equivalent*

a) *The rows of A are linearly independent/spanning/a basis for k^n*

b) The columns of A are linearly independent/spanning/a basis for k^n
c) The determinant of A is non-zero
d) The equation $A\mathbf{x} = \mathbf{0}$ has a unique solution.
e) For each $\mathbf{b} \in k^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution
f) For each $\mathbf{b} \in k^n$, the equation $\mathbf{y}^T A = \mathbf{b}^T$ has a unique solution.
g) There exists an n-by-n matrix B such that $AB = I$
h) There exists an n-by-n matrix B such that $BA = I$
i) A is invertible.

Proof. □

5. Block matrices

Definition 5.1. Let m_1, m_2 and n_1, n_2 be positive integers. Let $M = m_1 + m_2$ and $N = n_1 + n_2$. Let A_{11} be an m_1-by-n_1 matrix, let A_{12} be an m_1-by-n_2 matrix, A_{21} be an m_2-by-n_1 matrix, and A_{22} be an m_2-by-n_2 matrix. Then the matrix

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right)$$

is a 2-by-2 block matrix.

Proposition 5.2. *We have*

$$\det\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det(A)\det(C)$$

Proof. If A is not invertible, then both sides of the equation are 0. If A is invertible,

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} I & A^{-1}B \\ 0 & C \end{pmatrix}$$

But it is clear that the determinant of $\begin{pmatrix} A & \\ & I \end{pmatrix}$ is $\det(A)^{-1}$ and the determinant of $\begin{pmatrix} I & A^{-1}B \\ 0 & C \end{pmatrix}$ is $\det(C)$, So we conclude the required factorization. □

Definition 5.3. In general, let $m_1, m_2, \ldots, m_k$ be positive integers and $n_1, n_2, \ldots, n_\ell$ be positive integers and let A_{ij} be an m_i-by-n_j matrix for each $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, \ell$. The resulting

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1\ell} \\ \hline A_{21} & A_{22} & \cdots & A_{2\ell} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline A_{k1} & A_{k2} & \cdots & A_{k\ell} \end{array}\right)$$

is a k-by-ℓ block matrix (it is also a M-by-N matrix where $M = m_1 + m_2 + \cdots + m_k$ and $N = n_1 + n_2 + \cdots + n_\ell$.)

Try to think about under what conditions we can multiply two block matrices.

Proposition 5.4. *Suppose $T : V \to W$ is a linear transformation. Suppose that $W = W_1 \oplus W_2 \oplus \cdots \oplus W_m$ and $V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$*

Then there exists $T_i : V \to W_i$ such that for $\mathbf{v} \in V$, $T(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v}) + \cdots + T_m(\mathbf{v})$.

Furthermore, if $B_1, \ldots, B_n$ are bases for $V_1, \ldots, V_n$ respectively, and $C_1, \ldots, C_m$ are bases for $W_1, \ldots, W_m$ respectively, and $A_{ij} = {}_{C_i}[T_i|_{V_j}]_{B_j}$ then the block matrix of T is:

$$\left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1n} \\ \hline A_{21} & A_{22} & \cdots & A_{2n} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline A_{m1} & A_{m2} & \cdots & A_{mn} \end{array}\right)$$

Proof. Omitted. □

Definition 5.5. Let $A_1, \ldots, A_r$ be matrices such that A_i is an n_i-by-n_i matrix. Let $N = n_1 + n_2 + \cdots + n_r$ and define the N-by-N matrix

$$A_1 \oplus A_2 \oplus \cdots \oplus A_r = \left(\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_n \end{array}\right)$$

6. Polynomials

Definition 6.1. Let k be a field. A polynomial with coefficients in k is a sum

$$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$$

where $c_0, c_1, \ldots, c_n \in k$.

Theorem 6.2. *The polynomial ring $k[x]$ is an infinite dimensional vector space over k with basis $x^0, x^1, x^2, \ldots, x^n, \ldots$.*

Proof. The ring $k[x]$ is constructed to have this property. Here is a construction of $k[x]$:

a) Consider the set of sequences $f : \mathbb{N} \to k$ (certainly a vector space)
b) Let $k[x]$ be the subset of those f such that $f_i \neq 0$ for finitely many i
c) identify $1 \leftrightarrow (1, 0, 0, \ldots)$, $x \leftrightarrow (0, 1, 0, \ldots)$ and so on
d) define the natural addition, and scalar multiplication (as well as multiplication)

□

Definition 6.3. A monic polynomial is a polynomial of the form $x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ where $n \geq 0$. In other words, the leading coefficient of a monic polynomial is 1.

Definition 6.4. Let $f(x) = c_n x^n + \cdots + c_1 x + c_0$ with $c_n \neq 0$. Then define $\deg f(x) = n$.

Notice that $\deg fg = \deg f + \deg g$.

Definition 6.5. Let a, b be polynomials, and suppose $a \neq 0$. Then there exist unique polynomials q, r such that

$$b = qa + r$$

such that $\deg r < \deg a$ or $r = 0$.

The polynomial r is called the remainder and q is called the quotient.

Proof. By induction on $N = \deg a$. If $N = 0$, then a is a non-zero scalar. So $b = (b/a) \cdot a + 0$ as required.

Otherwise write $a = cx^N + a'$ with $\deg a' \leq N - 1$ or $a' = 0$. If $\deg b < \deg a$ or $b = 0$, then we may take $q = 0$ and $r = b$.

Suppose $\deg b \geq \deg a$. Write $b = dx^{N+j} +$ lower order terms. Then let $b' = b - (d/c)x^j \cdot a$. Notice that $b' = dx^{N+j} +$ lower order terms of $b - dx^{N+j} - (d/c)x^j a'$.

If $a' = 0$, then let By induction we may find q', r' such that $b' = q'a' + r'$ with $\deg r' < \deg a'$ □

Definition 6.6. Let I be a subspace of $k[x]$. Then I is called an ideal of $k[x]$ if $fI \subset I$ for all $f \in k[x]$.

Theorem 6.7. *Let I be an ideal of $k[x]$. Suppose $I \neq 0$. Then there exists a monic polynomial $m(x)$ such that*

$$I = \{f(x) \cdot m(x) \mid f(x) \in k[x]\}.$$

Proof. Suppose $I \neq 0$. Let f be a monic polynomial of smallest degree such that $f \in I$. (here we are using the well-ordering principle, that every non-empty subset of natural numbers has a least element).

We claim that every element of I is a multiple of f.

Let $g \in I$. Write $g = qf + r$ with $\deg r < \deg f$, or $r = 0$. We have that $g \in I$ and $qf \in I$ as well since $f \in I$.

So $r = g - qf \in I$. But then $r = 0$ (since otherwise r is an element of I with smaller degree than f).

So $g = q \cdot f$, as required. □

Definition 6.8. Let $f(x), g(x) \in k[x]$. Suppose $f(x)$ and $g(x)$ are not both zero. Then the greatest common divisor of $f(x)$ and $g(x)$, denoted $\gcd(f, g)$, is defined to be the monic polynomial $h(x)$ with the largest degree such that $f(x) = q(x)h(x)$ and $g(x) = r(x)h(x)$.

Theorem 6.9. *Let $f, g \in k[x]$ not both zero. Let $h = \gcd(f, g)$. Then there exists polynomials a, b such that*

$$h(x) = a(x)f(x) + b(x)g(x).$$

Proof. Let $I = \{a \cdot f + b \cdot g \mid a, b \in k[x]\}$. Check that I is an ideal of $k[x]$. Let $h(x)$ be the monic generator of I. Since $h \in I$, we have $h = af + bg$ for some $a, b \in k[x]$.

We now will establish that h is the gcd of f, g.

Notice that $f = 1 \cdot f + 0 \cdot g \in I$ so $f = qh$ for some q. Similarly, $g(x) = r(x)h(x)$ for some $r(x)$.

Suppose H is another polynomial which divides both f, g. Then H divides also $af + bg = h$, which implies $\deg H \leq \deg h$ as required. □

Definition 6.10. The polynomial $f(x)$ is called reducible if $f(x) = g(x)h(x)$ and g, h are both non-constant polynomials.

A non-constant polynomial $f(x)$ is called irreducible if it is not reducible.

Theorem 6.11. *Let $f(x) \in k[x]$ be a non-constant polynomial. Then there are monic, irreducible polynomials $p_1, \ldots, p_n$ such that*

$$f(x) = cp_1 \cdots p_n$$

where c is the leading coefficient of $f(x)$.

Proof. Sketch.

a) If f is irreducible we are done
b) else write $f = gh$ each having degree strictly less than the degree of f
c) find a factorization of g, h

d) for uniqueness, prove that if p is irreducible and p divides gh then p divides g or p divides h

□

Theorem 6.12. *The only irreducible polynomials in* $\mathbb{C}[x]$ *are* $x - a$ *for* $a \in \mathbb{C}$*.*

Proof. $\mathbb{C}$ is algebraically closed. So if $f(x) \in \mathbb{C}[x]$ is a non-constant polynomial, f has a root in $\mathbb{C}$. So if $\deg f > 1$ it is not reducible since we can write $f(x) = (x - a)g(x)$ where $f(a) = 0$.

The only irreducible monic polynomials are $x - a$ where $a \in \mathbb{C}$ □

7. Algebra of linear transformations

Definition 7.1. Let V and W be vector spaces. The set of all linear transformations from V to W is denoted by $\mathcal{L}(V, W)$.

Theorem 7.2. *The set of all linear transformations from* V *to* W *is a vector space. The addition and scalar multiplication are defined as follows:*

$$(T + S)(\mathbf{v}) = T(\mathbf{v}) + S(\mathbf{v}) \text{ for all } \mathbf{v} \in V$$

for all $T, S \in \mathcal{L}(V, W)$ *and*

$$(cT)(\mathbf{v}) = c \cdot (T(\mathbf{v})) \text{ for all } \mathbf{v} \in V$$

Proof. Verify that $T+S$ and cT are both linear transformations. Verify that the zero function $0 : V \to W$ defined by $0(\mathbf{v}) = \mathbf{0}_W$ for all $\mathbf{v} \in V$ is a linear transformation.

Let $\mathcal{F}$ be the set of all functions from $V \to W$. We claim that $\mathcal{F}$ is a vector space under the above operations. Then $\mathcal{L}(V, W)$ will be a vector space by the subspace test.

THe zero vector of $\mathcal{F}$ is the zero function $0(\mathbf{v}) = \mathbf{0}_W$ for all $\mathbf{v} \in V$.

The negative of $g \in \mathcal{F}$ is the function $-g$ defined by $(-g)(\mathbf{v}) = -(g(\mathbf{v}))$ for all $\mathbf{v} \in V$.

The other six axioms can be checked (it is kind of tedious though). If you have never tried it, then please try. □

Definition 7.3. Let V be a vector space. The set of all linear transformations $T : V \to V$ is denoted $\mathcal{L}(V)$.

We know this is a vector space over k. But in fact it is also known to be something called a "k-algebra". Other examples of k-algebras are the polynomial ring $k[x]$, and the n-by-n matrices over k.

Definition 7.4. Suppose that $\mathcal{A}$ is a k-vector space. If, in addition, there is a product

$$\mathcal{A} \times \mathcal{A} \to \mathcal{A}$$
$$(A, B) \mapsto A \cdot B$$

such that

a) $A \cdot (\beta B + C) = \beta AB + AC$ for all $A, B, C \in \mathcal{A}$ and $c \in k$
b) $(\alpha A + B) \cdot C = \alpha AC + BC$ for all $A, B, C \in \mathcal{A}$ and $c \in k$
c) there exists $1 \in \mathcal{A}$ such that $1 \cdot A = A \cdot 1 = A$ for all $A \in \mathcal{A}$
d) $A(BC) = (AB)C$ for all $A, B, C \in \mathcal{A}$.

then we say that $\mathcal{A}$ is a k-algebra.

Theorem 7.5. *The space $\mathcal{L}(V)$ is a k-algebra.*

Proof. Omitted. □

Theorem 7.6. *Let V be a vector space with basis B, and W a vector space with basis B'. Suppose $n = \dim(V)$ and $m = \dim(W)$. Then define a map*

$$\Psi : \mathcal{L}(V, W) \to k^{m \times n}$$

by the rule

$$\Psi(T) =_{B'} [T]_B$$

for all $T \in \mathcal{L}(V, W)$.

Then Ψ is an isomorphism of vector spaces.

Now, suppose that $V = W$ and $B = B'$ so that

$$\Psi(T) =_B [T]_B$$

for all $T : V \to V$, then

$$\Psi : \mathcal{L}(V) \to k^{n \times n}$$

is an isomorphism of k-algebras (in particular, $\Psi(c \cdot I_V) = c \cdot I_n$ and $\Psi(AB) = \Psi(A)\Psi(B)$).

Proof. Try it. □

8. Inner Product Spaces

In this chapter especially, we will take $k = \mathbb{C}$ or $k = \mathbb{R}$. Our prototype inner product for real vector spaces is the dot product $\langle \mathbf{v}, \mathbf{u} \rangle = \sum_{i=1}^{n} v_i u_i = \mathbf{v}^T \mathbf{u}$. In general, we have the following definition.

Definition 8.1. Let V be a real vector space. A symmetric inner product on V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

such that

a) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{v}, \mathbf{w} \in V$.
b) $\langle a\mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = a\langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a \in \mathbb{R}$
c) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v} \in V$ with equality if and only if $\mathbf{v} = \mathbf{0}$.

If $\langle \cdot, \cdot \rangle$ is a symmetric inner product on V then $(V, \langle \cdot, \cdot \rangle)$ is called a real inner product space.

We can define a similar notion for complex vector spaces, but we have to be careful with the symmetry. The prototypical complex inner product space is $\mathbb{C}^n$ with inner product $\langle \mathbf{v}, \mathbf{u} \rangle = \sum_{i=1}^n \overline{v_i} u_i = \mathbf{v}^* \mathbf{u}$ where $\mathbf{v}^*$ means the conjugate transpose of $\mathbf{v}$. In general, we have the following definition.

Definition 8.2. Let V be a complex vector space. A hermitian inner product on V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$$

such that

a) $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$ for all $\mathbf{v}, \mathbf{w} \in V$.
b) $\langle a\mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = \overline{a}\langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a \in \mathbb{C}$
c) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v} \in V$ with equality if and only if $\mathbf{v} = \mathbf{0}$.

If $\langle \cdot, \cdot \rangle$ is an hermitian inner product on V then $(V, \langle \cdot, \cdot \rangle)$ is called a complex inner product space.

Remark 8.3. From now on, whenever we say "inner product space", unless otherwise noted, we mean "(real or complex) inner product space".

Definition 8.4. Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in an inner product space. We say that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\langle \mathbf{v}, \mathbf{u} \rangle = 0$.

If S, T are non-empty subsets of V, then we say that S and T are orthogonal if $\langle \mathbf{s}, \mathbf{t} \rangle = 0$ for all $\mathbf{s} \in S$ and $\mathbf{t} \in T$.

Proposition 8.5. *Let V be a (real or complex) inner product space. Then*

a) $\langle \mathbf{0}, \mathbf{v} \rangle = 0$ *for all* $\mathbf{v} \in V$
b) *if* $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ *for all* $\mathbf{v} \in V$ *then* $\mathbf{u} = \mathbf{0}$.

Proof. Omitted. □

Definition 8.6. Let V be a (real or complex) inner product space. For each $\mathbf{v} \in V$, define the norm of $\mathbf{v}$, denoted by $\|\mathbf{v}\|$, by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

Proposition 8.7. *Let V be an inner product space. Then*

a) $\|vecv\| \geq 0$ *with equality if and only if* $\mathbf{v} = \mathbf{0}$
b) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$
c) $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle$
d) *if* $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ *then* $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$
e) $\|\mathbf{v} + \mathbf{u}\|^2 + \|\mathbf{v} - \mathbf{u}\|^2 = 2\|\mathbf{v}\|^2 + 2\|\mathbf{u}\|^2$

Proof. Omitted. □

Definition 8.8. Let $\mathbf{u}$ be a non-zero vector. Let $\mathbf{v} \in V$. We define the projection of $\mathbf{v}$ onto $\mathbf{u}$ to be the vector

$$\mathbf{x} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}\mathbf{u} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$$

Proposition 8.9. *Let $\mathbf{u}$ be a non-zero vector. Let $\mathbf{x} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}\mathbf{u}$ be the projection of $\mathbf{v}$ onto $\mathbf{u}$. Then $\mathbf{v} = \mathbf{x} + (\mathbf{v} - \mathbf{x})$ is a decomposition of $\mathbf{v}$ into two orthogonal vectors.*

Proof. Omitted. □

Theorem 8.10. *Let V be a (real or complex) inner product space. Then*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\|\|\mathbf{w}\|$$

with equality if and only if $\mathbf{v}, \mathbf{w}$ are linearly dependent.

Proof. The theorem is true when either $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$. Now, write $\mathbf{v} = \mathbf{x} + (\mathbf{v} - \mathbf{x})$ as in Proposition 8.9. Then since $\mathbf{x}$ is orthogonal to $\mathbf{v} - \mathbf{x}$, we have

$$\begin{aligned} \|\mathbf{v}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{v} - \mathbf{x}\|^2 \geq \|\mathbf{x}\|^2 \\ &= \|\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{w}\|^2}\mathbf{w}\| \\ &= \frac{|\langle \mathbf{v}, \mathbf{w} \rangle|^2}{\|\mathbf{w}\|^2} \end{aligned}$$

Now, if the above inequality is an equality then $\mathbf{x} = \mathbf{v}$ so $\mathbf{w}$ is in the span of $\mathbf{v}$.

Conversely, if $\mathbf{w} = c\mathbf{v}$ check that $|\langle \mathbf{w}, \mathbf{v} \rangle| = |c|\|\mathbf{v}\|^2 = \|\mathbf{v}\|\|\mathbf{w}\|$. □

Corollary 8.11. *We have*

$$\|\mathbf{v}+\mathbf{u}\| \le \|\mathbf{v}\| + \|\mathbf{u}\|$$

Proof. Omitted. □

Definition 8.12. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be non-zero vectors. The list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is called mutually orthogonal if

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \text{ for all } 1 \le i, j \le n \text{ with } i \ne j.$$

If the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is mutually orthogonal, it is called orthonormal if in addition $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ for all $i = 1, 2, \ldots, n$.

If the list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of orthonormal vectors is such that $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is called an orthonormal basis for V.

Proposition 8.13. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a list of non-zero, orthonormal vectors. Then if $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, we have*

$$c_1 = \langle \mathbf{v}_1, \mathbf{v} \rangle,\ c_2 = \langle \mathbf{v}_2, \mathbf{v} \rangle, \ldots,\ c_n = \langle \mathbf{v}_n, \mathbf{v} \rangle.$$

In particular, a list of orthonormal vectors is linearly independent.

Proof. Consider

$$\begin{aligned}
\langle \mathbf{v}_1, \mathbf{v} \rangle &= \langle \mathbf{v}_1, c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \rangle \\
&= c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + c_2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle + \cdots + c_n\langle \mathbf{v}_1, \mathbf{v}_n \rangle \\
&= c_1
\end{aligned}$$

since $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 1$ and $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$ and so on. The calculations for $c_2, c_3, \ldots, c_n$ are similar. □

Definition 8.14. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a list of orthonormal vectors and $U = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. The orthogonal projection of a vector $\mathbf{v}$ onto U is defined to be $P_U(\mathbf{v})$ where

$$P_U(\mathbf{v}) = \sum_{i=1}^{n} \langle \mathbf{u}_i, \mathbf{v} \rangle \mathbf{u}_i.$$

The orthogonal projection $P_U : V \to V$ satisfies

$$P_U^2 = P_U$$

and

$$P_U(V) = U$$

Theorem 8.15. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for V. Then there orthornormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ such that*

a) $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_i) = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_i)$ *for each* $i = 1, 2, \ldots, n$

b) *Letting* $U_i = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_i)$*, set* $\mathbf{y}_{i+1} = \mathbf{v}_{i+1} - P_{U_i}(\mathbf{v}_{i+1})$ *and then* $\mathbf{u}_{i+1} = \frac{\mathbf{y}_{i+1}}{\|\mathbf{y}_{i+1}\|}$

Proof. Omitted. $\square$

Definition 8.16. A linear functional on V is a linear transformation $f : V \to k$ where k is the field of scalars of V.

The set of all linear functionals on V is a vector space, called the dual of V, denoted $V^\vee$ or $\mathcal{L}(V, k)$.

If V is an inner product space and $\mathbf{u}$ is a fixed vector, then we can construct a linear functional by defining $f(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$.

Theorem 8.17. *Let V be a finite dimensional inner product space. There is a 1-1 correspondence between linear functionals and vectors of V.*

Proof. For $\mathbf{v} \in V$ define $f_v : V \to k$ by the rule

$$f_{\mathbf{v}}(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle.$$

THe properties of inner product spaces tell us that $f_{\mathbf{v}}$ is linear.

Furthermore, $f_{\mathbf{v}+c\mathbf{w}} = f_{\mathbf{v}} + c f_{\mathbf{w}}$. So the map $\mathbf{v} \mapsto f_{\mathbf{v}}$ is a linear transformation

$$V \to V^\vee.$$

Let $f : V \to k$ be a linear transformation (in other words, $f \in V^\vee$). If $f(\mathbf{w}) = 0$ for all $\mathbf{w}$ then $f = f_{\mathbf{0}}$.

Else there exists $\mathbf{w} \in V$ such that $f(\mathbf{w}) \neq 0$. By the rank-nullity theorem (Theorem 1.8), if U is the nullspace of f, U has dimension $n - 1$. Let $\mathbf{u}$ be a unit vector spanning the orthogonal complement of U. Then calculate $c = f(\mathbf{u})$ and notice that $f = f_{c\mathbf{u}}$. $\square$

Remark 8.18. There is a Riesz Representation Theorem for complete inner product spaces (a.k.a. Hilbert spaces).

Proposition 8.19. *Let V, W be two inner product spaces with orthonormal bases B for V and B' for W. Let $T : V \to W$. Then the matrix of T relative to B and B' is given by the matrix with i, j-th entry $\langle \mathbf{w}_i, T(\mathbf{v}_j) \rangle$.*

Proof. To compute the matrix of T, we compute the coordinates of $T(\mathbf{v}_j)$ relative to B', the i-th coordinate is given by

$$\langle \mathbf{w}_i, T(\mathbf{v}_j) \rangle$$

by Proposition 8.13. $\square$

Definition 8.20. Let $T : V \to W$ be a linear transformation. A linear transformation $S : W \to V$ is called an adjoint of T if

$$\langle \mathbf{w}, T(\mathbf{v})\rangle = \langle T^*(\mathbf{w}), \mathbf{v}\rangle$$

for all $\mathbf{w} \in W$ and $\mathbf{v} \in V$.

The matrix of the adjoint is the conjugate transpose of the matrix.

Definition 8.21. Let $T : V \to V$ be a linear transformation and V are inner product spaces. Then T is called an isometry if

$$\|T(\mathbf{v})\| = \|\mathbf{v}\|$$

for all $\mathbf{v} \in V$.

Proposition 8.22. *Let $T : V \to V$ and $S : V \to V$ be isometries. Then ST is an isometry. T is 1-1. If V is finite-dimensional then T is invertible and T^{-1} is an isometry.*

Proof. Notice

$$\|ST(\mathbf{v})\| = \|S(T(\mathbf{v}))\| = \|T(\mathbf{v})\| = \|\mathbf{v}\|$$

since both S and T are isometries.

If $T(\mathbf{v}) = \mathbf{0}$ then $0 = \|T(\mathbf{v})\| = \|\mathbf{v}\|$ so $\mathbf{v} = \mathbf{0}$. So T is 1-1.

By the rank-nullity theorem, if T is 1-1 and V is finite-dimensional, then T is invertible. Writing $T^{-1}\mathbf{w} = \mathbf{v}$ if and only if $T(\mathbf{v}) = \mathbf{w}$ we have

$$\|\mathbf{v}\| = \|T(\mathbf{v})\|$$

so

$$\|T^{-1}(\mathbf{w})\| = \|\mathbf{w}\|$$

□

Proposition 8.23. *Let $T : V \to V$ be a linear transformation. Then $\langle T(\mathbf{v}), T(\mathbf{u})\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$ for all $\mathbf{v}, \mathbf{u} \in V$ if and only if T is an isometry.*

Proof. Suppose T is an isometry. There is a trick to proving that $\langle T(\mathbf{v}), T(\mathbf{u})\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$. It is to expand

$$\langle T(\mathbf{v} \pm \mathbf{u}), T(\mathbf{v} \pm \mathbf{u})\rangle$$

and compare to $\langle \mathbf{v} \pm \mathbf{u}, \mathbf{v} \pm \mathbf{u}\rangle$.

Of course, if $\langle T(\mathbf{v}), T(\mathbf{u})\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$ then T is an isometry because you can put $\mathbf{v} = \mathbf{u}$. □

Proposition 8.24. *Let $T : V \to V$ where V is finite-dimensional. Then T is an isometry if and only if $T^*T = I$.*

Proof. We have

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, T^*T(\mathbf{v}) \rangle$$

Subtracting $\langle \mathbf{u}, \mathbf{v} \rangle$ from both sides:

$$0 = \langle \mathbf{u}, T^*T\mathbf{v} - \mathbf{v} \rangle$$

for all $\mathbf{v}, \mathbf{u} \in V$. Therefore, $T^*T\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$ so $T^*T = I$.

If $T^*T = I$ then

$$\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, T^*T\mathbf{u} \rangle = \langle T\mathbf{v}, T\mathbf{u} \rangle$$

proving that T is unitary. □

Definition 8.25. Let U be an n-by-n matrix. Then U is called unitary if $U^*U = I$.

By Proposition 8.24 and Proposition 8.19, a unitary matrix is the matrix representation of an isometry relative to an orthonormal basis, and it is also an isometry $k^n \to k^n$.

A unitary matrix with real entries, is called an orthogonal matrix. Real orthogonal matrices satisfy $Q^TQ = I$.

Proposition 8.26. *Let U, V be n-by-n matrices and W and m-by-m matrix. THen*

a) If U, V are unitary then so is UV
b) If U and W is unitary then so is

$$U \oplus W = \begin{pmatrix} U & 0 \\ 0 & W \end{pmatrix}$$

c) If U is unitary then $|\det(U)| = 1$.

Proof. Omitted. □

Theorem 8.27. *Let U be an n-by-n matrix. The following are equivalent.*

a) U is unitary
b) U^T, U^* are unitary
c) the columns of U form an orthonormal basis for k^n (remember, $k = \mathbb{R}$ or $\mathbb{C}$ depending on if we have a real or complex inner product space)
d) the rows of U form an orthonormal basis

Proof. Omitted. □

Definition 8.28. A rank 1 projection matrix is a matrix $P = \mathbf{u}\mathbf{u}^*$ where $\mathbf{u}$ is a non-zero unit vector (in other words, $\|\mathbf{u}\| = 1$.

We have that $P_U = P$ for $U = \operatorname{span}(\mathbf{u})$ in the terminology of Theorem 8.15.

Proposition 8.29. *Let P be a rank 1 projection matrix, corresponding to unit vector $\mathbf{u}$.*

a) *the columnspace of P is* $\operatorname{span}(\mathbf{u})$
b) *If $\langle \mathbf{v}, \mathbf{u}\rangle = 0$ then $P_U(\mathbf{v}) = \mathbf{0}$*
c) *If $\mathbf{v} = c\mathbf{u}$ then $P_U(\mathbf{v}) = \mathbf{v}$.*
d) $P_U^* = P_U$
e) $P_U^2 = P_U$

Proof. We have $P(\mathbf{v}) = \mathbf{u}\mathbf{u}^*\mathbf{v} = \mathbf{u}\langle \mathbf{u}, \mathbf{v}\rangle$ so the columnspace (in other words, the range of P) is equal to the span of $\mathbf{u}$.

For part b), use the above formula again.

For part c) $P(c\mathbf{u}) = \mathbf{u}\mathbf{u}^*(c\mathbf{u}) = c\mathbf{u}$

Compute $P_U^* = (\mathbf{u}\mathbf{u}^*)^* = \mathbf{u}\mathbf{u}^*$

Similarly, $P_U^2 = \mathbf{u}\mathbf{u}^*\mathbf{u}\mathbf{u}^* = \mathbf{u}\mathbf{u}^*$ since $\mathbf{u}^*\mathbf{u} = 1$. □

Definition 8.30. Let $\mathbf{w} \neq \mathbf{0}$. Let $\mathbf{u} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ with rank 1 projection

$$P_{\mathbf{u}} = \mathbf{u}\mathbf{u}^* = \frac{\mathbf{w}\mathbf{w}^*}{\mathbf{w}^*\mathbf{w}}$$

The Householder matrix corresponding to $\mathbf{w}$ is defined to be

$$U_{\mathbf{w}} = 1 - 2P_{\mathbf{u}}$$

with corresponding Householder transformation

$$\mathbf{x} \mapsto \mathbf{x} - 2\langle \mathbf{x}, \mathbf{u}\rangle \mathbf{u}$$

Theorem 8.31. *Let U be a Householder matrix. Then $U^* = U = U^{-1}$, If U is a real Householder matrix then $U^T = U = U^{-1}$.*

Proof.

$$U^*U = (1 - 2P_{\mathbf{u}})^2 = 1 - 4P_{\mathbf{u}} + 4P_{\mathbf{u}}^2 = 1$$

So U is unitary, $U^* = U$ since $P_{\mathbf{u}}^* = P_{\mathbf{u}}$, and so

$$U = U^* = U^{-1}.$$

If U is real then $U^T = U^*$ so the second part follows. □

Theorem 8.32. *Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$. Suppose $0 \neq \|\mathbf{x}\| = \|\mathbf{y}\|$. Let*

$$\sigma = \begin{cases} 1 & \text{if } \langle \mathbf{x}, \mathbf{y}\rangle \leq 0 \\ -1 & \text{if } \langle \mathbf{x}, \mathbf{y}\rangle > 0 \end{cases},$$

and let $\mathbf{w} = \mathbf{y} - \sigma\mathbf{x}$. Then $\sigma U_{\mathbf{w}}$ is real orthogonal and $\sigma U_{\mathbf{w}}\mathbf{x} = \mathbf{y}$.

Proof. We just need to check that $\sigma U_{\mathbf{w}}\mathbf{x} = \mathbf{y}$. Let $\mathbf{w}_+ = \mathbf{x} - \mathbf{y}$ and $\mathbf{w}_- = \mathbf{x} + \mathbf{y}$. Then notice that $\langle \mathbf{w}_+, \mathbf{w}_-\rangle = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 0$.

So

$$\begin{aligned}
U_{\mathbf{w}_-}\mathbf{w}_+ &= \mathbf{w}_+ \\
U_{\mathbf{w}_-}\mathbf{w}_- &= -\mathbf{w}_- \\
U_{\mathbf{w}_+}\mathbf{w}_+ &= -\mathbf{w}_+ \\
U_{\mathbf{w}_-}\mathbf{w}_- &= \mathbf{w}_-
\end{aligned}$$

Therefore, writing $\mathbf{x} = (1/2)\mathbf{w}_+ + (1/2)\mathbf{w}_-$,

$$U_{\mathbf{w}_-}\mathbf{x} = (1/2)\mathbf{w}_+ - (1/2)\mathbf{w}_- = \mathbf{y},$$

and similarly,

$$U_{\mathbf{w}_+}\mathbf{x} = -\mathbf{y}$$

Accounting for the definition of σ and $\mathbf{w}$ proves the theorem. □

Theorem 8.33. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ *and suppose* $\|\mathbf{x}\| = \|\mathbf{y}\| \neq 0$. *Let*

$$\sigma = \begin{cases} 1 & \text{if } \langle \mathbf{x}, \mathbf{y} \rangle = 0, \\ -\overline{\langle \mathbf{x}, \mathbf{y} \rangle}/|\langle \mathbf{x}, \mathbf{y} \rangle| & \text{if } \langle \mathbf{x}, \mathbf{y} \rangle \neq 0, \end{cases},$$

and let $\mathbf{w} = \mathbf{y} - \sigma\mathbf{x}$. *Then* $\sigma U_{\mathbf{w}}$ *is unitary and* $\sigma U_{\mathbf{w}}\mathbf{x} = \mathbf{y}$.

Proof. Omitted (but similar to the real case). □

Theorem 8.34. *Let A be an m-by-n matrix and suppose that $m \geq n$. There exists an m-by-m unitary matrix V and upper triangular n-by-n matrix R whose diagonal entries are real and non-negative, such that*

$$A = V \begin{pmatrix} R \\ 0 \end{pmatrix}.$$

If $V = (Q\ Q')$ in which Q contains the first n columns of V, then Q has orthonormal columns and $A = QR$.

If $\operatorname{rank}(A) = n$, *then the factors Q and R are unique and R has positive diagonal entries.*

Proof. Let $\mathbf{a}_1$ be the first column of A. Let $c = \|\mathbf{a}_1\|$. Use a Householder matrix U_1 such that

$$U_1 A = \begin{pmatrix} c & \cdot \\ \mathbf{0} & A' \end{pmatrix}$$

where A' is an $m-1$-by-$n-1$ matrix.

Roughly, we then apply induction. □

9. Diagonalization and the Cayley-Hamilton Theorem

Definition 9.1. Let A be an n-by-n matrix. Then λ is called an eigenvalue for A if there exists $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$ such that

$$A\mathbf{v} = \lambda \mathbf{v}.$$

If $\mathbf{v} \neq \mathbf{0}$ and $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{C}$, then $\mathbf{v}$ is called an eigenvector for A.

The pair $(\lambda, \mathbf{v})$ such that $A\mathbf{v} = \lambda\mathbf{v}$ and $\mathbf{v} \neq \mathbf{0}$ is called an eigenpair.

Theorem 9.2. *Let A be an n-by-n matrix. Let $\lambda \in \mathbb{C}$. The following are equivalent*

a) λ *is an eigenvalue for* A
b) λ *is an eigenvalue for* A^T
c) $A\mathbf{v} = \lambda\mathbf{v}$ *for some* $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$
d) $(A - \lambda I)\mathbf{v} = \mathbf{0}$ *has a non-trivial solution*
e) $A - \lambda$ *is not invertible*
f) $A^T - \lambda$ *is not invertible*

Proof. Omitted. □

Definition 9.3. Let A be an n-by-n matrix. Then $p_A(z) = \det(zI - A) = z^n + c_{n-1}z^{n-1} + \cdots + c_1 z + c_0$. Then p_A is a monic polynomial of degree n, called the characteristic polynomial.

Each coefficient of p_A is a polynomial in the entries of A, $c_{n-1} = -\operatorname{tr}(A)$ and $c_0 = (-1)^n \det A$.

Proof. Omitted. □

Proposition 9.4. *The characteristic polynomial of*

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

is $p_A p_B$.

Proof. Compute

$$\det \begin{pmatrix} A - zI & B \\ 0 & C - zI \end{pmatrix} = \det(A - zI)\det(C - zI)$$

by Proposition 5.2. □

Definition 9.5. Let $k = \mathbb{C}$ or $\mathbb{R}$. We say a matrix A is diagonalizable over k if there is an invertible matrix P and a diagonal matrix D (with entries in k) such that

$$A = PDP^{-1}$$

We say that a complex matrix A is unitarily diagonalizable if there is a unitary matrix U and diagonal D such that

$$A = UDU^*$$

We say that a real matrix A is orthogonally diagonalizable if there is an orthogonal matrix Q and diagonal matrix D with real entries such that

$$A = QDQ^T.$$

Theorem 9.6. *Let $k = \mathbb{C}$ or $\mathbb{R}$.*

A matrix is diagonalizable over k if and only if there is a basis of k^n consisting of eigenvectors for A.

Proof. Suppose $A = PDP^{-1}$. Then $AP = PD$. The j-th column of the left hand side is

$$A\mathbf{v}_j$$

where $\mathbf{v}_j$ is the j-th column of P. The j-th column of the right hand side is $\lambda_j \mathbf{v}_j$ so each column of P is an eigenvector. Since P is invertible, there is a basis of k^n of eigenvectors.

Each argument may be reversed as well. □

Theorem 9.7. *Let A be an n-by-n matrix. Then there is a monic polynomial $p \in \mathbb{C}[x]$, with $\deg p \leq n^2$ such that $p(A) = 0$.*

Proof. The dimension of the vector space of n-by-n matrices is n^2. So $1, A, A^2, \ldots, A^{n^2}$ are linearly independent which is enough to guarantee the existence of such polynomial. □

Theorem 9.8. *Let A be an n-by-n matrix. Then A has an eigenvalue (in $\mathbb{C}$).*

Proof. Let $p(x)$ be a monic polynomial with $p(A) = 0$. Since $\mathbb{C}$ is algebraically closed, we can write

$$p(x) = (x - a_1)(x - a_2) \cdots (x - a_n).$$

So

$$(A - a_1)(A - a_2) \cdots (A - a_n) = 0.$$

If $A-a_1$ is not invertible then a_1 is an eigenvalue. If $A-a_1$ is invertible then

$$(A-a_2)\cdots(A-a_n)=0,$$

and we proceed similarly until we find that some a_j is an eigenvalue of A. □

Theorem 9.9. *Let A be an n-by-n matrix, with eigenpair $(\lambda, \mathbf{v})$ such that $\|\mathbf{v}\|=1$. Then there is a unitary matrix*

$$U=(\mathbf{v}\ U')$$

and an upper triangular matrix T such that

$$A=UTU^*$$

and $t_{11}=\lambda$ and $t_{11}, t_{22}, \ldots, t_{nn}$ are the eigenvalues of A.

Furthermore, if A is a real matrix with real eigenvalues $\lambda_1, \ldots, \lambda_n$ and $\mathbf{v}$ has real entries, then there exists an orthogonal matrix $Q=(\mathbf{v}Q')$ such that $A=QTQ^T$ (therefore, also T has real entries as well with $t_{ii}=\lambda_i$ for each i).

Proof. We will prove in the case that A is real. We proceed by induction on n. So suppose that every real matrix with real eigenvalues and real eigenvector $\mathbf{x}$ which is a unit vector, then we can write

$$A=QTQ^T$$

where the first column of Q is $\mathbf{x}$ and T is upper triangular.

Let A be a real matrix with real eigenvalues. Let $\mathbf{x}$ be an unit eigenvector. Then Theorem 8.32, there is a unitary matrix U with first column equal to $\mathbf{x}$ (U maps $\mathbf{e}_1$ to $\mathbf{x}$). Write $U=(\mathbf{x}\ U')$. Then

$$AU=(A\mathbf{x}\ AU')$$

Since the columns of U are orthonormal $U'^T\mathbf{x}=0$,so

$$U^TAU=\begin{pmatrix}\mathbf{x}^T\\ U'^T\end{pmatrix}(\lambda_1\mathbf{x}\ AU')=\begin{pmatrix}\lambda_1 & \mathbf{x}^TAU'\\ \lambda_1U'^T\mathbf{x} & U'^TAU'\end{pmatrix}=\begin{pmatrix}\lambda_1 & \\ \mathbf{0} & U'^TAU'\end{pmatrix}.$$

By induction $A'=U'^TAU'$ can also be written as VTV^T.

The eigenvalues of A are $\lambda_1, \ldots, \lambda_n$ so the eigenvalues of A' must be $\lambda_2, \ldots, \lambda_n$

Now, let $V_1=1\oplus V$ and let $U_1=V_1U$ is a unitary matrix, and a computation confirms that $U_1^TAU_1$ is an upper-triangular matrix. □

Theorem 9.10. *Let A be an n-by-n matrix and $p(x)$ its characteristic polynoimal. Then $p(A)=0$.*

Proof. Omitted. □

Definition 9.11. Let A be an n-by-n matrix. Then A is called normal if $AA^* = A^*A$.

Theorem 9.12. *Let A be an n-by-n matrix. The following are equivalent.*

a) A is normal (Definition 9.11)
b) A is unitarily diagonalizable (Definition 9.5)
c) $\mathbb{C}^n$ has an orthonormal basis consisting of eigenvectors of A

Now, let A be a real n-by-n matrix. The following are equivalent:

a) A is symmetric
b) A is real orthogonally diagonalizable (there exists an orthogonal matrix Q such that $A = QDQ^T$ for some diagonal D)
c) $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A

Proof. Omitted. □

10. Canonical forms

Definition 10.1. Let $T : V \to V$ be a linear operator. Let U be a subspace of V. We say that U is T-invariant if $T(U) \subset U$.

If U is T-invariant, then the restriction of T to U, $T|_U$, is a linear operator on U.

Proposition 10.2. *Suppose $T : V \to V$ is a linear transformation. Suppose $V = U \oplus W$ and U is T-invariant. Let $B = \mathbf{u}_1, \ldots, \mathbf{u}_k$ be a basis for U and $B' = \mathbf{w}_1, \ldots, \mathbf{w}_\ell$ be a basis for W. Then $B \cup B'$ is a basis for V and the matrix of T relative to $B \cup B'$ (see Proposition 5.4)*

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right)$$

If W is also T-invariant, then the matrix of T is of the form

$$\left(\begin{array}{cc} A_{11} & 0 \\ \hline 0 & A_{22} \end{array}\right)$$

Proof. Let $C = B \cup B'$. The columns of the matrix of T correspond to

$$[T(\mathbf{u}_1)]_C,\ [T(\mathbf{u}_2)]_C,\ \ldots,\ [T(\mathbf{u}_k)]_C,\ [T(\mathbf{w}_1]_C, \ldots, [T(\mathbf{w}_\ell)]_C.$$

But for $\mathbf{v} \in V$, Proposition 3.2 or Proposition 5.4, we have that $\mathbf{v} = \mathbf{u} + \mathbf{w}$ and

$$[\mathbf{v}]_C = \left(\frac{[\mathbf{u}]_B}{[\mathbf{w}]_{B'}}\right).$$

Now, it remains to note that for $\mathbf{u}_j \in B$, $T(\mathbf{u}_i) \in U$, so $T(\mathbf{u}_j) = \mathbf{u} + \mathbf{0}$ according to Theorem 2.11. This implies that the matrix of T is of the form

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right).$$

If W is also T-invariant, a similar argument applies. In this case, our notation Definition 5.5 applies and we write the matrix of T as

$$A_{11} \oplus A_{22}$$

where A_{11} is the matrix of $T|_U : U \to U$ and A_{22} is the matrix of $T|_W : W \to W$. □

Proposition 10.3. *Let $p(x) \in k[x]$ be a polynomial. Let $T : V \to V$ be a linear operator. Let U be a T-invariant subspace. Then $p(T)(U)$ is T-invariant. Also $(p(T))^{-1}(U) = \{\mathbf{v} \in V \mid p(T)(\mathbf{v}) \in U\}$ is T-invariant.*

In particular, $\ker(p(T))$ *and* $p(T)(V)$ *are T-invariant subspaces.*

Proof. Omitted. □

Proposition 10.4. *Let $W = \ker(T - \lambda)$ and let $U \subseteq W$. Then U is a T-invariant subspace.*

Proof. Omitted. □

Lemma 10.5. *Suppose $T : V \to V$ is a linear operator. Suppose $f(T) = 0$. Suppose $f(x) = g(x)h(x)$ and* $\gcd(g, h) = 1$*. Then*

$$V = \ker(g(T)) \oplus \ker(h(T)).$$

Proof. Since $\gcd(g, h) = 1$ write $1 = ag + bh$ for some polynomials $a, b \in k[x]$ (Theorem 6.9). Let $\mathbf{v} \in V$. Then write

$$\mathbf{v} = 1\mathbf{v} = a(T)g(T)\mathbf{v} + b(T)h(T)\mathbf{v}$$

and let $\mathbf{v}_1 = b(T)h(T)\mathbf{v}$ and let $\mathbf{v}_2 = a(T)g(T)\mathbf{v}$. Notice that $\mathbf{v}_1 \in \ker(g(T))$ and $\mathbf{v}_2 \in \ker(h(T))$. So $V = \ker(g(T)) + \ker(h(T))$ but we have to show that the sum is a direct sum (Definition 2.1)

So suppose $\mathbf{v} \in \ker(g(T)) \cap \ker(h(T))$. Then write

$$\mathbf{v} = a(T)g(T)\mathbf{v} + b(T)h(T)\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

since $g(T)\mathbf{v} = h(T)\mathbf{v} = \mathbf{0}$.

So the sum is a direct sum as required. □

Theorem 10.6. *Let $p_T(z) = (z - \lambda_1)^{m_1} \cdots (z - \lambda_k)^{m_k}$ be the factorization of p_T over the complex numbers. Then*

$$V = \ker(T - \lambda_1)^{m_1} \oplus \ker(T - \lambda_2)^{m_2} \oplus \cdots \oplus \ker(T - \lambda_k)^{m_k}$$

Proof. Let $f_1(z) = (z - \lambda_2)^{m_2} \cdots (z - \lambda_k)^{m_k}$ and let $V_1 = \ker(f_1(T))$. Then $\gcd((x - \lambda_1)^{m_1}, (x - \lambda_2)^{m_2}, \dots, (x - \lambda_k)^{m_k}) = 1$. Therefore, by Lemma 10.5,

$$V = \ker(T - \lambda_1)^{m_1} \oplus \ker(f_1(T)).$$

Now, let $T_1 : V_1 \to V_1$ be the restriction of T to V_1.

To apply induction we need for $f_1(z)$ to be the characteristic polynomial of T_1. We have $V_1 = (T - \lambda_1)^{m_1}(V)$. The matrix of T is

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

Now, A is an m_1-by-m_1 matrix whose only eigenvalue is λ_1. So the characteristic polynomial of A is $(x - \lambda_1)^{m_1}$. The characteristic polynomial of $A \oplus B$ is the product of the characteristic polynomials of A and B. So the characteristic polynomial of B must be $f_1(z)$. By induction,

$$V_1 = \ker(T - \lambda_2)^{m_2} \oplus \cdots \oplus \ker(T - \lambda_k)^{m_k}$$

and so

$$V = \ker(T - \lambda_1)^{m_1} \oplus V_1 = \ker(T - \lambda_1)^{m_1} \oplus \cdots \oplus \ker(T - \lambda_k)^{m_k}$$

add reference for characteristic polynomial of direct sum of matrices

$\square$

Definition 10.7. A matrix A is called *nilpotent* if $A^n = 0$ for some $n \geq 1$.

Proposition 10.8. *Let A be a square matrix. Then* $\operatorname{Spec} A = \{\lambda\}$ *if and only if $A - \lambda$ is nilpotent.*

Proof. Let $B = A - \lambda$.

If $B^n = 0$ and $B\mathbf{v} = \lambda\mathbf{v}$ then $\mathbf{0} = B^n\mathbf{v} = \lambda^n\mathbf{v}$ so $\lambda = 0$.

On the other hand, if $\operatorname{Spec} B = \{0\}$, then the characteristic polynomial of B must be x^n and then apply the Cayley-Hamilton Theorem (Theorem 9.10). $\square$

Theorem 10.9. *Let A be an n-by-n matrix, with characteristic polynomial $p(x)$ and minimal polynomial $m(x)$. The following are equivalent.*

a) *A is nilpotent ($A^k = 0$ for some $k \geq 1$)*
b) *$p(x) = x^n$*
c) *$m(x) = x^j$ for some $1 \leq j \leq n$*
d) *A has no non-zero eigenvalues*
e) *$A^n = 0$*

Proof. Omitted. □

Definition 10.10. Let $\lambda \in \mathbb{C}$ and $k \geq 1$. A Jordan block of size k with eigenvalue λ is

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & 1 \end{pmatrix}$$

We will now work with nilpotent operators for a little bit.

Definition 10.11. Let $T : V \to V$. Let $\mathbf{v} \in V$ and suppose that $T^k(\mathbf{v}) = \mathbf{0}$ and $T^{k-1}(\mathbf{v}) \neq \mathbf{0}$. Then the subspace $U = \operatorname{span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ is called a cyclic subspace of V. The vector $\mathbf{v}$ is called a cyclic vector. We will write $U = C(\mathbf{v})$ to mean that U is a cyclic subspace with cyclic vector $\mathbf{v}$.

Proposition 10.12. *Suppose $U = C(\mathbf{v})$. Then $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ is a basis for U where $T^m(\mathbf{v}) = \mathbf{0}$ and $T^{m-1}(\mathbf{v}) \neq \mathbf{0}$.*

Proof. Let $B = \mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$. By definition of cyclic subspace, B spans U. Suppose for some $0 \leq j \leq m-1$, we have $c_j \neq 0$ and

$$c_j T^j \mathbf{v} + c_{j+1} T^{j+1}(\mathbf{v}) + \cdots + c_{m-1} T^{m-1}(\mathbf{v}) = \mathbf{0}.$$

Then apply T^{m-j-1} to the equation, and using that $T^m \mathbf{v} = \mathbf{0}$, we have

$$c_j T^{m-1} \mathbf{v} = \mathbf{0}$$

and since $T^{m-1}\mathbf{v} \neq \mathbf{0}$, we conclude that $c_j = 0$. This is a contradiction. This argument tells us that if

$$c_0 + c_1 T\mathbf{v} + \cdots + c_{m-1} T^{m-1} \mathbf{v} = \mathbf{0}$$

then in the above expression $c_0 = 0$, $c_1 = 0$, and so on. So B is linearly indendent and so a basis for U. □

Proposition 10.13. *Suppose $T : V \to V$ is nilpotent. Then V is the direct sum of cyclic subspaces.*

Proof. By induction on $\dim V$ ($\dim V = 1$ is clear).

Suppose that the theorem is true for all W with $\dim W < \dim V$.

Let $W = T(V)$. THen $\dim W < \dim V$ by the rank-nullity theorem (since T is nilpotent, its nullspace is non-trivial).

So write $W = C(\mathbf{w}_1) \oplus \cdots \oplus C(\mathbf{w}_n)$.

Then write $T(\mathbf{v}_i) = \mathbf{w}_i$. Let $W' = C(\mathbf{v}_1) + \cdots + C(\mathbf{v}_n)$. We claim that W' is the direct sum

$$W' = C(\mathbf{v}_1) \oplus \cdots \oplus C(\mathbf{v}_n).$$

Consider

$$p_1(T)\mathbf{v}_1 + \cdots + p_n(T)\mathbf{v}_n = \mathbf{0}$$

We must show that $p_i(T)\mathbf{v}_i = \mathbf{0}$ for all $1 \leq i \leq n$.

First, suppose that $p_i(0) \neq 0$ for some i. Then $\gcd(p_i, x^{m_i}) = 1$, so there exists $a, b \in k[x]$ such that

$$ap_i + bx^{m_i} = 1$$

Now

$$\mathbf{v}_i = (a(T)p_i(T) + b(T)T^{m_i})\mathbf{v}_i = a(T)p_i(T)\mathbf{v}_i$$

in particular

$$\mathbf{v}_i = \sum_{j \neq i} a(T)p_j(T)\mathbf{v}_j$$

proving that $\mathbf{v}_i = \mathbf{0}$ which is a contradiction since then $\mathbf{w}_i = T(\mathbf{v}_i) = \mathbf{0}$ so the sum for $W = T(V)$ is not direct. Therefore, $p_i(0) = 0$ for all i. Therefore $p_i(x) = xq_i(x)$ for some polynomials q_i. In particular, each vector $p_i(T)\mathbf{v}_i = q_i(T)\mathbf{w}_i$ and the sum for W is a direct sum proving that $q_i(T)\mathbf{w}_i = \mathbf{0}$, so the sum for W' is a direct sum.

Finally, $W' + \ker(T) = V$. So find $U \subset \ker(T)$ such that $W' \oplus U = V$ and finish the proof by noting that any subspace of $\ker(T)$ is a direct sum of cyclic subspaces. □

Definition 10.14. A Jordan matrix is defined to be a direct sum of Jordan blocks:

$$J = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_k}(\lambda_k)$$

Definition 10.15. A nilpotent Jordan block is a matrix

$$J_n = J_n(0) = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Definition 10.16. A nilpotent Jordan matrix is a direct sum of nilpotent Jordan blocks

$$J_{n_1} \oplus J_{n_2} \oplus \cdots \oplus J_{n_k} = \begin{pmatrix} J_{n_1} & 0 & \cdots & 0 \\ 0 & J_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_k} \end{pmatrix}$$

Lemma 10.17. *Let* $J_n = (\mathbf{0}\ \mathbf{e}_1\ \cdots\ \mathbf{e}_{n-1})$. *Let* $1 \leq p \leq n-1$. *Then* $J_n^p = (\mathbf{0}\ \cdots \mathbf{e}_1\ \mathbf{e}_2\ \cdots \mathbf{e}_{n-p})$. *And* $J_n^n = 0$.

In particular, $\operatorname{rank}(J_n^p) = n - p$ *for all* $p \leq n$.
In particular, $\operatorname{rank}(J_n^p) - \operatorname{rank}(J_n^{p-1}) = 1$ *for* $p \leq n$ *and* 0 *for* $p > n$.

Proof. Notice that $J_n \mathbf{e}_{i+1} = \mathbf{e}_i$ for $i \leq n-1$. □

Theorem 10.18. *Let V be a finite-dimensional vector space. Let $T : V \to V$ be a linear operator. Suppose that $p_T(z) = (z-\lambda_1)^{m_1} \cdots (z-\lambda_n)^{m_n}$ is the characteristic polynomial of T. Then the Jordan normal form of T exists and is unique.*

Proof. Apply Theorem 10.6. Then apply Proposition 10.13. That proves existence.
Uniqueness is an exercise. □

11. Singular value decomposition and applications

Definition 11.1. Let P be a symmetric (if P is real) or Hermitian (if P has complex entries) n-by-n matrix. Let the eigenvalues of P be $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then P is called positive semi-definite if $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, ..., $\lambda_n \geq 0$.

The symmetric matrix P is called positive definite if all the eigenvalues are positive.

Proposition 11.2. *Let P be a (symmetric or Hermitian) n-by-n matrix. The following are equivalent*

a) *P is positive semi-definite*
b) $\mathbf{x}^* P \mathbf{x} \geq 0$ *for all* $\mathbf{x} \in \mathbb{C}^n$ *(if P is Hermitian)or for all* $\mathbf{x} \in \mathbb{R}^n$ *if P is symmetric.*

Proof. Omitted. □

Definition 11.3. Let P be a positive semi-definite matrix. Then there exists a unitary matrix U such that

$$P = UDU^*$$

and

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are all non-negative. Define the square root of P, denoted $P^{1/2}$, by

$$P^{1/2} = U \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{pmatrix} U^*$$

Definition 11.4. Let A be an m-by-n complex matrix. Let $r = \operatorname{rank}(A)$. Let $q = \min(m, n)$. Then A^*A is a positive semi-definite n-by-n matrix, with r positive eigenvalues. Then let the positive eigenvalues of $(A^*A)^{1/2}$ be $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ and define

$$\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_q = 0.$$

Then the singular values of A are defined to be

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_q.$$

Theorem 11.5. *Let A be an m-by-n matrix, let $r = \operatorname{rank}(A)$, let $q = \min(m, n)$ and let*

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_q$$

be the singular values of A and let $c \in \mathbb{C}$. Then

a) *$\sigma_1^2, \dots, \sigma_r^2$ are the positive eigenvalues of A^*A and AA^**
b) *$\sum_{i=1}^{q} = \sigma_i^2 = \operatorname{tr} A^*A = \operatorname{tr} AA^*$*
c) *A, A^*, A^T, and $\overline{A}$ have the same singular values*
d) *The singular values of cA are $|c|\sigma_1, |c|\sigma_2, \dots, |c|\sigma_q$.*

Proof. It is clear that the positive eigenvalues of A^*A are the squares of the positive eigenvalues of $(A^*A)^{1/2}$. The non-zero eigenvalues of AA^* and A^*A are the same.

$$\operatorname{tr} A^*A = \sum_{i=1}^{r} \sigma_1^r$$

and $\operatorname{tr} A^*A = \operatorname{tr} AA^*$ by cyclicity of trace.

The non-zero eigenvalues of A^*A and AA^* are the same. But

$$\overline{A^*A} = A^T\overline{A}, \overline{AA^*} = \overline{A}A^T$$

which means $A, \overline{A}, A^T, A^*$ have the same singular values. □

Theorem 11.6. *Let A be a non-zero m-by-n matrix Let $r = \operatorname{rank}(A)$. Let $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$ be the non-zero singular values of A. Define*

$$\Sigma_r = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

Then there is an n-by-n unitary matrix V and m-by-m unitary matrix W such that

$$A = V\Sigma W^*$$

in which

$$\Sigma = \begin{pmatrix} \Sigma_r & 0_{r\times n-r} \\ 0_{m-r\times r} & 0_{m-r\times n-r} \end{pmatrix}$$

is the same size as A.

Proof. Suppose $m \geq n$. Write $A^*A = WDW^*$ with unitary W. Then $D^{1/2} = \Sigma_r \oplus 0_{n-r}$ Let

$$E = \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

so that $D^{1/2}E^{-1} = I_r \oplus 0_{n-r}$.

Now, let $B = AWE^{-1}$ and consider

$$\begin{aligned} B^*B &= (AWE^{-1})^*(AWE^{-1}) \\ &= (E^{-1})^*W^*A^*AWE^{-1} \\ &= E^{-1}W^*WDW^*WE^{-1} \\ &= E^{-1}D^{1/2}D^{1/2}E^{-1} \\ &= I_r \oplus 0_{n-r} \end{aligned}$$

Write $B = (V_r\ B')$ so that V_r is the first r columns of B. NOtice $B^*B = \begin{pmatrix} V_r^*V_r & V_r^*B' \\ B'^*V_r & (B')^*B' \end{pmatrix} = I_r \oplus 0_{n-r}$.

So the columns of V_r are orthonormal, so they may be extended to an orthonormal basis of $\mathbb{C}^m$, so let $V = (V_r\ V')$ be a unitary matrix. On the other hand, $(B')^*B' = 0$ means that each columns of B' is zero, so B' is zero.

Now, let us compare AW and $V\Sigma$.

$$V\sigma = (V_r\ V')\begin{pmatrix} \Sigma_r & 0_{r\times n-r} \\ 0_{m-r\times r} & 0_{m-r\times n-r} \end{pmatrix} = (V_r\Sigma_r\ 0_{m\times n-r}).$$

$$AW = BE = (V_r\ 0)(\Sigma_r \oplus I_{n-r}) = (V_r\Sigma_r\ 0_{m\times n-r}$$

as required. □

12. Quadric surfaces

Definition 12.1. A quadric surface is a surface in $\mathbb{R}^3$ with an equation of the form

$$ax^2 + bxy + cxz + dy^2 + eyz + fz^2 + gx + hy + iz + l = 0,$$

where $a, b, c, d, e, f, g, h, i \in \mathbb{R}$ and at least one of a, b, c, d, e, f is non-zero. (A quadric surface is just a surface defined by a degree 2 equation in x, y, z).

Definition 12.2. A quadratic form (for our purposes) is a function $q : \mathbb{R}^n \to \mathbb{R}$ such that $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for a symmetric matrix A.

Definition 12.3. Let $S = \{ax^2 + bxy + \cdots + iz + l = 0\}$ be a quadric surface. Define a quadratic form $q_S(x, y, z, w) = ax^2 + bxy + cxz + dy^2 + eyz + fz^2 + gxw + hyw + izw + lw^2$ with associated matrix

$$A_S = \begin{pmatrix} a & b/2 & c/2 & g/2 \\ b/2 & d & e/2 & h/2 \\ c/2 & e/2 & f & i/2 \\ g/2 & h/2 & i/2 & l \end{pmatrix}$$

Definition 12.4. Let S be a quadric surface, with quadratic form q_S and matrix A_S. Then $A_S = QDQ^T$ since A_S is symmetric. Let $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ be the eigenvalues of A_S.

Then if one or more $\lambda_i = 0$ then S is called degenerate.

The goal is now to classify all the quadric surfaces. We will do this in class if we have time.

References

[Axl15] Sheldon Axler. *Linear algebra done right*. Third. Undergraduate Texts in Mathematics. Springer, Cham, 2015, pp. xviii+340. URL: https://doi.org/10.1007/978-3-319-11080-6.

[GH17] Stephan Ramon Garcia and Roger A. Horn. *A second course in linear algebra*. English. Cambridge: Cambridge University Press, 2017, pp. xix + 426.