Math 110: Linear Algebra Notes

1. **Vector Spaces**

   (a) A **field** F is a set on which two operations (addition and multiplication) are defined so that, for each pair $x, y \in F$, there are unique elements $x+y$ and $x \cdot y \in F$ for which the following conditions hold: commutativity of addition and multiplication, associativity of addition and multiplication, existence of identity elements (0 and 1) for addition and multiplication respectively, existence of inverses for addition and multiplication, and distributivity of multiplication over addition. Thus $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields while $\mathbb{N}, \mathbb{Z}$ are not.

   (b) A **vector space** V over a field F is a set on which two operations (addition and scalar multiplication) are defined such that the set is closed under the operations, and the following conditions hold for every $\vec{u}, \vec{v}, \vec{z} \in V$ and $a, b \in F$:

      i. Commutativity of addition: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
      ii. Associativity of addition: $(\vec{u} + \vec{v}) + \vec{z} = \vec{u} + (\vec{v} + \vec{z})$.
      iii. Additive identity: There is $\vec{0} \in V$ such that for all $\vec{u}$, $\vec{0} + \vec{u} = \vec{u}$.
      iv. Existence of inverse: For every $\vec{u}$, there is an element $\vec{w}$ such that $\vec{u} + \vec{w} = \vec{0}$.
      v. Multiplicative identity: There is $1\vec{u} = \vec{u}$.
      vi. Associativity of scalar multiplication: $(ab)\vec{u} = a(b\vec{u})$.
      vii. Distributivity of vectors: $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$.
      viii. Distributivity of scalars: $(a + b)\vec{u} = a\vec{u} + b\vec{u}$.

      The elements of F are **scalars** and the elements of V are **vectors**.

   (c) Examples of vector spaces:

      - F is a 1-dimensional vector space over itself.
      - The set of n-tuples with entries from a field F, denoted F^n.
      - The set of all $m \times n$ matrices with entries from a field F, denoted $M_{m\times n}(F)$.
      - The set of all polynomials of degree at most n with coefficients from a field F, denoted $\mathbb{P}_n(F)$.
      - The set of all functions from a non-empty set S to some field F, $\mathcal{F}(S, F)$ is a vector space with addition and scalar multiplication for $f, g \in \mathcal{F}(S, F)$ and $c \in F$ defined as $(f+g)(s) = f(s)+g(s)$ and $(cf)(s) = c[f(s)]$.

   (d) Some consequences of the definition of a vector space:

      - $\vec{x} + \vec{z} = \vec{y} + \vec{z} \implies \vec{x} = \vec{y}$
      - The additive inverse of $\vec{x}$ is given by $(-1)\vec{x}$.
      - The zero vector and the additive inverse are unique.

   (e) The following holds of scalar multiplication:

   $$0\vec{x} = \vec{0} \quad \forall x \in V \qquad a\vec{0} = \vec{0} \quad \forall a \in F$$

2. **Vector Subspaces**

   (a) A subset W of a vector space V over a field F is a **subspace** of V if W is a vector space over F with the operations of addition and scalar multiplication defined in V.

   (b) Theorem: W is a subspace of V iff the following conditions hold:

      - $\vec{0} \in W$
      - $\vec{x} + \vec{y} \in W \quad \forall \vec{x}, \vec{y} \in W$
      - $c\vec{x} \in W \qquad \forall c \in F, \vec{x} \in W$

      The other axioms are inherited from the vector space V, or follow from the above three. Note that $\vec{0}$ is the same zero vector from V.

   (c) Examples of subspaces:

      - For any vector space V, $\{\vec{0}\}$ and V itself are trivial subspaces.
      - Set of all continuous real-valued even functions on $\mathbb{R}$, subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$
        Recall: A function g is even if $g(-t) = g(t), \forall t \in F$ and odd if $g(-t) = -g(t), \forall t \in F$.
      - Set of all $n \times n$ matrices with trace equal to zero, subspace of $M_{n\times n}(F)$.

   (d) Matrix terminology:
   The **transpose** of a $m \times n$ matrix A is the $n \times m$ matrix A^T where $A^T_{ij} = A_{ji}$ for $1 \le i \le m$ and $1 \le j \le n$. A **symmetric** matrix is a matrix A such that $A = A^T$. We have$(aA + bB)^T = aA^T + bB^T$. A **diagonal matrix** has all non-diagonal entries zero. The **trace** of a square matrix is the sum of the diagonal entries. A **skew symmetric** matrix is a matrix A such that $A^T = -A$.

   (e) Constructions of subspaces

      i. The intersection of subspaces of a vector space V is a subspace of V.
      ii. The union of subspaces of a vector space V is *not* a subspace of V (unless one subspace contains the other).

   (f) Let W_1 and W_2 be subspaces of a vector space V. The **sum** of W_1 and W_2, denoted $W_1 + W_2$, is the set $\{\vec{x} + \vec{y} \mid \vec{x} \in W_1, \vec{y} \in W_2\}$. The sum is the smallest subspace of V that contains W_1 and W_2.

   (g) The vector space W is the **direct sum** of W_1 and W_2, denoted $W_1 \oplus W_2$, if $W = W_1 + W_2$ and $W_1 \cap W_2 = \{\vec{0}\}$. As a consequence, every vector in W has a unique decomposition $\vec{w_1} + \vec{w_2}$ for $\vec{w_1} \in W_1$, $\vec{w_2} \in W_2$. Examples:

- For any matrix A, we have $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$, where $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric.
- For any function f, $f = \frac{1}{2}(f(t) + f(-t)) + \frac{1}{2}(f(t) - f(-t))$, where $f(t) + f(-t)$ is even and $f(t) - f(-t)$ is odd.

(h) Two sets S_1 and S_2 are **equal** iff $S_1 \subseteq S_2$ and $S_2 \subseteq S_1$. That is, every element in each set is in the other set.

3. **Linear Combinations and Systems of Equations**

(a) A vector $\vec{v} \in V_F$ is a **linear combination** of vectors $\vec{u_1}, \vec{u_2}, \ldots, \vec{u_n}$ if $\vec{v} = a_1\vec{u_1} + a_2\vec{u_2} + \ldots + a_n\vec{u_n}$ for scalars $a_i \in F$.

(b) Solving a system of equations amounts to finding the $a_1, a_2, \ldots, a_n$ scalars needed to express a certain vector as a linear combination of others. Three operations can be used to simplify a system without changing the set of solutions:

i. Interchanging the order of equations
ii. Multiplying any equation by a nonzero constant
iii. Adding a constant multiple of any equation to another equation

(c) Let S be a nonempty subset of V. The **span** of S, denoted $\text{Span}(S)$, is the set consisting of all linear combinations of the vectors in S. Note $\text{Span}(\emptyset) = \{\vec{0}\}$.

(d) Theorem: The span of any subset S of V is a subspace of V. Further, any subspace that contains S must also contain $\text{Span}(S)$.

(e) A subset S of a vector space V **generates/spans** V if $\text{Span}(S) = V$.

(f) If S_1 and S_2 are subsets of V, then

- $\text{Span}(S_1 \cup S_2) = \text{Span}(S_1) + \text{Span}(S_2)$
- $\text{Span}(S_1 \cap S_2) \subseteq \text{Span}(S_1) \cap \text{Span}(S_2)$

4. **Linear Dependence and Independence**

(a) A subset S of a vector space V is **linearly dependent** if there exists a finite number of distinct vectors $\vec{u_1}, \vec{u_2}, \ldots, \vec{u_n} \in S$ and scalars $a_1, a_2, \ldots, a_n$, not all zero, such that

$$a_1\vec{u_1} + a_2\vec{u_2} + \ldots + a_n\vec{u_n} = \vec{0}$$

That is, at least one vector can be written as a linear combination of the others. Consequently, any subset that contains $\vec{0}$ is linearly dependent.

(b) A subset S of a vector space V is **linearly independent** if whenever we have

$$c_1\vec{u_1} + c_2\vec{u_2} + \ldots + c_n\vec{u_n} = \vec{0}$$

then we must have $c_1 = c_2 = \ldots = c_n = 0$ for $\vec{u_1}, \vec{u_2}, \ldots, \vec{u_n} \in S$. Hence, the empty set and any set consisting of a single nonzero vector is linearly independent.

5. **Basis and Dimension**

(a) A **basis** β for a vector space V is a linearly independent subset of V that generates V. Examples:

- $\emptyset$ is a basis for the zero vector space $\{\vec{0}\}$.
- Let $\vec{e_1} = (1, 0, 0, \ldots, 0)$, $\vec{e_2} = (0, 1, 0, \ldots, 0)$, $\vec{e_n} = (0, 0, 0, \ldots, 1)$. Then $\{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}\}$ is the standard basis for F^n.
- The set $\{1, x, x^2, \ldots, x^n\}$ is the standard basis for $\mathbb{P}_n(F)$.

(b) The **dimension** of a vector space V, denoted $\dim(V)$, is the number of vectors in a basis for V.

(c) Replacement Theorem: Let V be a vector space generated by a set G containing n vectors, and L be a linearly independent subset of V containing m vectors. Then $m \leq n$ and there exists a subset H of G with $n - m$ vectors such that $L \cup H$ generates V.

(d) Consequences:

i. If V is finite-dimensional, then every basis for V contains the same number of vectors.
ii. Any linearly independent subset of V that contains exactly $n = \dim(V)$ vectors is a basis for V.
iii. Every linearly independent subset of V can be extended into a basis for V.

(e) Dimension examples:

- F^n has dimension n.
- $M_{m\times n}(F)$ has dimension mn.
- $\mathbb{P}_n(F)$ has dimension $n + 1$.
- $\mathbb{C}$ has dimension 1 over $\mathbb{C}$ and dimension 2 over $\mathbb{R}$.
- A basis for the set of all symmetric $n \times n$ matrices is $\{A^{ij} \mid 1 \leq i \leq j \leq n\}$ where A^{ij} is the $n \times n$ matrix with 1 in the ith row and j^{th} column, 1 in the jth row and i^{th} column, and 0 elsewhere. The dimension is $\frac{1}{2}n(n + 1)$.
- The dimension of the set of all $n \times n$ skew-symmetric matrices is $\frac{1}{2}n(n-1)$ since we must have 0s on the diagonal.

(f) Dimensions of subspaces: Let W be a subspace of a finite-dimensional vector space V. Then $\dim(W) \leq \dim(V)$. If $\dim(W) = \dim(V)$, then $W = V$.

(g) For any two subspaces W_1 and W_2 of a finite-dimensional vector space V,

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$$

6. **Linear Transformations**

(a) Let V and W be vector spaces over F. A function $T : V \mapsto W$ is a **linear transformation** if $\forall \vec{x}, \vec{y} \in V, c \in F$, we have

i. $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$

ii. $T(c\vec{x}) = cT(\vec{x})$

(b) Properties of T:

- If T is linear, $T(\vec{0}_V) = \vec{0}_W$
- T is linear iff T is closed under linear combination.

(c) In $\mathbb{R}^n$, rotation, reflection, and projection are examples of linear transformations. Differentiation and integration are linear transformations on vector spaces of continuous functions.

(d) Let $T : V \mapsto W$ be a linear transformation.

i. The **null space** $\mathcal{N}(T)$ (or **kernel** Ker T) of T is $\{\vec{x} \in V \mid T(\vec{x}) = \vec{0}_W\}$.

ii. The **range** (or **image** Im T) is the subset of W consisting of all images of vectors in V; that is, $\{T(\vec{x}) \mid \vec{x} \in V\}$.

(e) The kernel and image of T are subspaces of V and W, respectively.

(f) If $\beta = \{\vec{v_1}, \ldots, \vec{v_n}\}$ is a basis for V, then

$$\text{Im } T = \text{Span}\left(\{T(\vec{v_1}), \ldots, T(\vec{v_n})\}\right)$$

(g) Define the **nullity** and **rank** of T to be the dimensions of the null space and range, respectively.

(h) **Dimension Theorem**: Let V and W be vector spaces, and let $T : V \mapsto W$ be linear. If V is finite-dimensional, then

$$\text{nullity}(T) + \text{rank}(T) = \dim(V)$$

(i) A function $f : A \mapsto B$ is

- **injective** (**one-to-one**) if whenever $f(x) = f(y)$, then $x = y$. Equivalently, $f(x) = 0$ iff $x = 0$.
- **surjective** (**onto**) if $B = \text{Im } f$.

(j) T is one-to-one iff $\mathcal{N}(T) = \{\vec{0}\}$. T is onto iff Im $(T) = W$.

(k) If V and W are vector spaces of equal and finite dimension, and $T : V \mapsto W$ is linear, then the following are equivalent: T is onto, T is one-to-one, and $\text{rank}(T) = \dim(V)$.

(l) A linear transformation is completely determined by its action on a basis. That is, if $\{\vec{v_1}, \ldots, \vec{v_n}\}$ is a basis for V and $T : V \mapsto W$ is linear, then for any $\vec{w_1}, \ldots, \vec{w_n} \in W$, there exists (exactly one) linear transformation T such that $T(\vec{v_i}) = \vec{w_i}$ for $1 \leq i \leq n$.

7. **Matrix Representation of a Transformation**

(a) There is a one-to-one correspondence between linear transformations and matrices that allows us to utilize properties of one to study properties of the other.

(b) Let $\beta = \{\vec{u_1}, \vec{u_2}, \ldots, \vec{u_n}\}$ be an *ordered* basis for V. For $\vec{x} \in V$, let $a_1, a_2, \ldots, a_n$ be the unique scalars such that $x = \sum_{i=1}^{n} a_i \vec{u_i}$. Then the **coordinate vector of $\vec{x}$ with respect to** β, denoted $[\vec{x}]_\beta$, is

$$[\vec{x}]_\beta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

(c) Suppose V and W are finite vector spaces with ordered bases $\beta = \{\vec{v_1}, \ldots, \vec{v_n}\}$ and $\gamma = \{\vec{w_1}, \ldots, \vec{w_m}\}$, respectively. Let $T : V \mapsto W$ be linear. The **matrix of T with respect to bases β and γ**, denoted $[T]_\beta^\gamma$, is the $m \times n$ matrix

$$[T]_\beta^\gamma = \begin{bmatrix} | & | & & | \\ [T(\vec{v_1})]_\gamma & [T(\vec{v_2})]_\gamma & \cdots & [T(\vec{v_n})]_\gamma \\ | & | & & | \end{bmatrix}$$

(d) Let V and W be vectors spaces over F. Let $T, U : V \mapsto W$ be linear.

- Using the usual definition of addition and scalar multiplication for functions, $\forall a \in F$, $aT + U$ is linear. Further, $[aT + U]_\beta^\gamma = a[T]_\beta^\gamma + [U]_\beta^\gamma$.
- The set of all linear transformations from V to W, denoted $\mathcal{L}(V, W)$, is a vector space over F.

(e) Interesting result: If $\dim(V) = \dim(W)$ and $T : V \mapsto W$ is linear, then there exists bases β of V and γ of W such that $[T]_\beta^\gamma$ is diagonal. To see this, take $\beta = \{\vec{b_1}, \ldots, \vec{b_r}, \vec{b_{r+1}}, \ldots, \vec{b_n}\}$ where $\vec{b_1}, \ldots, \vec{b_r}$ is a basis for Ker T, and $\vec{b_{r+1}}, \ldots, \vec{b_n}$ extend it to a basis of V. Then take $\gamma = \{\vec{c_1}, \ldots, \vec{c_r}, \vec{c_{r+1}}, \ldots, \vec{c_n}\}$ where $\vec{c_i} = T(\vec{b_i})$ for $i \geq r+1$ and $\vec{c_1}, \ldots, \vec{c_r}$ extend it to a basis of W. The $T(\vec{b_i})$ for $i \geq r+1$ are linearly independent because whenever we have $\sum a_i T(\vec{b_i}) = T(\sum a_i \vec{b_i}) = 0$, $\sum a_i \vec{b_i}$ must be in Ker T. By the construction of β, this can ony be true if $a_i = 0$. So $[T]_\beta^\gamma$ is diagonal whose diagonal has r zeroes and $n - r$ ones.

(f) The composition of transformations T and U, denoted TU, is linear and means $T(U(\vec{v}))$. With regards to the matrices of the transformations, this corresponds to matrix multiplication. Indeed, matrix multiplication was defined for this purpose.

(g) Matrix multiplication: Let A be a $m \times n$ matrix and B be a $n \times p$ matrix. AB is the $m \times p$ matrix such that AB_{ij} is the sum of the products of the corresponding entries from the ith row of A and

j^{th} column of B. Precisely,

$$AB_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \qquad \text{for } 1 \le i \le m,\ 1 \le j \le p$$

Properties:

i. $A(B+C) = AB + AC$
ii. $a(AB) = (aA)B = A(aB)$ for any scalar a
iii. $IA = A = AI$, where I is the identity matrix
iv. $A(BC) = (AB)C$
v. $(AB)^T = B^T A^T$

(h) Let V, W, Z be finite-dimensional vector spaces with ordered bases α, β, γ, respectively. Let $T : V \mapsto W$ and $U : W \mapsto Z$ be linear. Then

$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$$

(i) Let V and W be finite-dimensional vector spaces with ordered bases β, γ, respectively and let $T : V \mapsto W$ be linear. Then for any $\vec{v} \in V$,

$$[T(\vec{v})]_\gamma = [T]_\beta^\gamma [\vec{v}]_\beta$$

(j) Let A be an $m \times n$ matrix with entries from F. The left-multiplication transformation $L_A : F^n \mapsto F^m$ is linear. Further, $[L_A]_\beta^\gamma = A$.

(k) If UT is one-to-one, then T is one-to-one. If UT is onto, then U is onto. For both cases, this is illustrated by $T : F \mapsto F^2$ defined by $T(a) = (a, 0)$ and $U : F^2 \mapsto F$ defined by $U((a_1, a_2)) = a_1$.

(l) If $V = W$, that is, $T : V \mapsto V$, then T is called a **linear operator** on V. The matrix of T with respect to a basis β of V is denoted $[T]_\beta$. Such transformations will be the topic of study later.

8. **Invertibility and Isomorphisms**

(a) Let V and W be vector spaces, and $T : V \mapsto W$ be linear. A function $U : W \mapsto V$ is the **inverse** of T if $TU = I_W$ and $UT = I_V$. If T is invertible, this inverse is unique and is denoted by T^{-1}.

(b) A function T is invertible iff it is both one-to-one and onto. If V and W are of equal and finite dimension, these conditions are equivalent.

(c) Analogous to linear transformations, an $n \times n$ matrix A is invertible if there exists an $n \times n$ matrix B such that $AB = BA = I$. This matrix is unique and is denoted A^{-1}.

(d) Properties:

- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$
- $(A^{-1})^{-1} = A$

(e) Let V and W be finite-dimensional vector spaces with ordered bases β, γ, respectively and let $T : V \mapsto W$ be linear. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible. Furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

(f) Let V and W be vector spaces. V and W are **isomorphic** if there exists a linear transformation $T : V \mapsto W$ that is invertible. T is called an isomorphism from V onto W.

(g) V and W are isomorphic iff $\dim(V) = \dim(W)$. To see this, let $\beta = \{\vec{v_1}, \ldots, \vec{v_n}\}$ be a basis for V and $\gamma = \{\vec{w_1}, \ldots, \vec{w_n}\}$ be a basis for W, and define $T : V \mapsto W$ where $T(\vec{v_i}) = \vec{w_i}$. Then $[T]_\beta^\gamma$ is the identity matrix, so T is invertible.

(h) Now we formally show the correspondence between linear transofmations and matrcies: Let V with ordered basis β and W with ordered basis γ be vector spaces with dimension n and m, respectively. The function $\Phi : \mathcal{L}(V, W) \mapsto M_{m\times n}(F)$, defined by $\Phi(T) = [T]_\beta^\gamma$, is an isomorphism. Consequently, $\dim(\mathcal{L}(V, W)) = mn$.

9. **Change of Basis**

(a) Let β and γ be two ordered bases for a finite dimensional vector space V. The matrix $Q = [I_V]_\gamma^\beta$ is a **change of coordinate matrix** that changes γ-coordinates to β-coordinates.

(b) Q is invertible and for any $\vec{v} \in V$, $[\vec{v}]_\beta = Q[\vec{v}]_\gamma$.

(c) Theorem: Let T be a linear transformation on V, β and γ be ordered bases for V, and Q be the change of coordinate matrix from γ to β coordinates. Then

$$[T]_\gamma = Q^{-1}[T]_\beta Q$$

(d) Let A and B be matrices in $M_{n\times n}(F)$. We say B is **similar** to A, denoted $B \sim A$, if there exists an invertible matrix Q such that $B = Q^{-1}AQ$. This is an equivalence relation (reflexive, symmetric, and transitive), so we say A is similar to B.

(e) A linear transformation $T : V \mapsto F$ is called a **linear functional** on V.

(f) The **dual space** of V, denoted V^*, is the vector space $\mathcal{L}(V, F)$. Since $\dim(V^*) = \dim(V)$, V and V^* are isomorphic.

(g) The ordered basis $B^* = \{f_1, \ldots, f_n\}$ of V^*, where f_i is the ith coordinate function with respect to β, is called the **dual basis** of β.

10. **Elementary Matrices and Systems of Equations**

(a) The three elementary (row or column) operations on a matrix are (1) interchanging rows, (2) multiplying any row by a nonzero scalar, and (3) adding a scalar multiple of one row to another.

(b) An $n \times n$ **elementary matrix** is an invertible matrix obtained by performing an elementary operation on I_n. If A is an $m \times n$ matrix and E is an elementary $m \times m$ matrix, then EA is the matrix obtained from A by performing the same elementary row operation as that which produces E from I_m. (Same for column operations, except $n \times n$ matrix E, and AE).

(c) Every invertible matrix is a product of elementary matrices. This is justification behind the computation of an inverse: If A is an invertible $n \times n$ matrix, then $(A|I_n)$ can be transformed into $(I_n|A^{-1})$ by a finite number of elementary row operations.

(d) The image and kernel of linear transformations are analogous to the column space and null space, respectively, of matrices. Some important properties:

- $\text{rank}(A^T) = \text{rank}(A)$
- $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$

(e) A system of equations $A\vec{x} = \vec{0}$ is called **homogeneous**.

(f) Let s be any solution to $A\vec{x} = \vec{b}$. The solution set to $A\vec{x} = \vec{b}$ is the sum of the solution set to $A\vec{x} = \vec{0}$ and s. To see this, consider two solutions $\vec{s}$ and $\vec{t}$ to the system; then $\vec{t} - \vec{s}$ is in the null space of A.

(g) Let $A\vec{x} = \vec{b}$. A is invertible iff the system has exactly one solution.

(h) Two systems of linear equations are **equivalent** if they have the same solution set. Let $A\vec{x} = \vec{b}$ be a system of m linear equations in n unknowns, and C be an invertible $m \times m$ matrix. Then the system $(CA)\vec{x} = C\vec{b}$ is equivalent to $A\vec{x} = \vec{b}$. Hence row operations on a system do not change the solution set (while column operations do).

(i) A matrix is said to be in **reduced row echelon form** if the following conditions are satisfied:

i. Any row containing a nonzero entry precedes any row in which all entries are zero (if any).
ii. The first nonzero entry in each row is the only nonzero entry in its column.
iii. The first nonzero entry in each row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row.

(j) **Gaussian elimination** is the most efficient procedure to reduce an augmented matrix to reduced row echelon form.

11. **Determinants**

(a) If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a 2×2 matrix over F, the determinant of A, denoted $\det(A)$, is $ad - bc$.

(b) Geometric interpretation: The area of the parallelogram determined by $\vec{u}$ and $\vec{v}$, $\vec{u}, \vec{v} \in \mathbb{R}^2$, is

$$\left| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right|$$

(c) The determinant is *not* a linear transformation, but it is a linear function of each row when the other rows are held fixed.

(d) The determinant of a $n \times n$ matrix can be evaluated by **Cofactor expansion** along any row or column. For rows, for any $1 \leq i \leq n$, we have

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij})$$

where $\tilde{A}_{ij}$ denotes the matrix obtained by removing row i and column j from A. The term $(-1)^{i+j} \det(\tilde{A}_{ij})$ is called the ij-cofactor of A.

(e) Effects of elementary row (or column) operations on the determinant: If B is a matrix obtained by...

- interchanging any two rows of A, then $\det(B) = -\det(A)$.
- multiplying any row of A by a nonzero scalar k, then $\det(B) = k \det(A)$.
- adding a scalar multiple of one row of A to another row of A, then $\det(B) = \det(A)$.

(f) Results:

- $\det(kA) = k^n \det(A)$
- $\det(A^T) = \det(A)$
- The determinant of an upper triangular matrix is the product of its diagonal entries.
- A is invertible iff $\det(A) \neq 0$. Further, $\det(A^{-1}) = \frac{1}{\det(A)}$.

(g) For $A, B \in M_{n \times n}(F)$, $\det(AB) = \det(A)\det(B)$. The proof involves first showing the result if A is an elementary matrix. Then split into cases if A is invertible or not, using the fact that A is a product of elementary matrices if it is.

(h) A simple consequence is that if $A \sim B$, then $\det(A) = \det(B)$.

(i) Suppose $M \in M_{n \times n}(F)$ is of the form

$$M = \begin{bmatrix} A & B \\ O & C \end{bmatrix}$$

where A and C are square matrices and O is the zero matrix. Then $\det(M) = \det(A)\det(C)$.
Proof: Let

$$N = \begin{bmatrix} A & B \\ O & I \end{bmatrix}, \quad N' = \begin{bmatrix} I & O \\ O & C \end{bmatrix}$$

Then $M = N'N$. Since $\det(N) = \det(A)$ (apply cofactor expansion along bottom row), and $\det(N') = \det(C)$, the proof is complete.

(j) If A is an $n \times n$ matrix whose diagonal entries are p and all off-diagonal entries are q, $\det(A) = (p-q)^{n-1}(p + (n-1)q)$.

(k) If $A \in M_{n \times n}(\mathbb{R})$ and the rows of A are $a_1, \ldots, a_n$, then $|\det(A)|$ is the n-dimensional volume of the parallelepiped having vectors $a_1, \ldots, a_n$ as adjacent sides.

12. **Eigenvalues and Eigenvectors**

(a) The discussion in this section applies to matrices and linear operators interchangeably; consider $A = [T]_\beta$.

(b) Let T be a linear operator on V. A *nonzero* vector $\vec{v} \in V$ is an **eigenvector** of T if there exists a scalar λ such that $T(\vec{v}) = \lambda\vec{v}$. The scalar λ is called the **eigenvalue** corresponding to eigenvector $\vec{v}$.

(c) A linear operator T on a finite-dimensional vector space V is **diagonalizable** iff there is an ordered basis β for V consisting of eigenvectors of T. If so, $[T]_\beta$ is a diagonal matrix.

(d) Computation of Eigenvalues: Let $A \in M_{n\times n}(F)$. Then a scalar λ is an eigenvalue of A iff $\det(A - \lambda I_n) = 0$. *Proof:* λ is an eigenvalue of A iff there exists a nonzero vector $\vec{v} \in F^n$ such that $A\vec{v} = \lambda\vec{v}$, that is $(A - \lambda I_n)\vec{v} = \vec{0}$. This is true iff $A - \lambda I_n$ is not invertible, which is equivalent to the statement $\det(A - \lambda I_n) = 0$.

(e) The polynomial $\det(A - \lambda I_n)$, denoted $\mathrm{char}_A(\lambda)$, is called the **characteristic polynomial** of A. It is a polynomial of degree n with leading coefficient $(-1)^n$. The roots are eigenvalues, so A has at least 0 and at most n distinct eigenvalues.

(f) Note on computing the characteristic polynomial: recall that adding scalar multiples of rows or columns to each other does not affect the determinant. Use this to zero out as many matrix entries as possible.

(g) The second term of the characteristic polynomial is $(-1)^{n-1}\lambda^{n-1}\ \mathrm{tr}(A)$ and the last term is $\det(A)$.

(h) Properties:

- $\mathrm{char}_A(\lambda) = \mathrm{char}_{A^T}(\lambda)$
- If $A \sim B$, then $\mathrm{char}_A(\lambda) = \mathrm{char}_B(\lambda)$

(i) Computation of Eigenvectors: A nonzero vector $\vec{v} \in V$ is an eigenvector of A corresponding to λ iff $\vec{v} \in \mathcal{N}(A - \lambda I_n)$.

(j) Examples for transformations on $\mathbb{R}^2$: reflection across a line has eigenvalues 1 and -1, projection onto a line has eigenvalues 0 and 1, and rotation by θ has no (real) eigenvalues. The following matrix rotates a vector counterclockwise by θ:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

(k) Let $A \in M_{n\times n}(F)$

- A has n eigenvalues, *counting each according to their multiplicity.*
- The sum of the n eigenvalues of A is the same as the trace of A.
- The product of the n eigenvalues of A is the same as the determinant of A.

13. **Diagonalizability**

(a) A set of eigenvectors of T, each corresponding to a different eigenvalue of T, is linearly independent. (*Proof*: Proceed with induction on k, the number of eigenvectors. In the inductive step, apply $T - \lambda_{k+1}I$ to both sides of $a_1\vec{v_1} + \ldots + a_{k+1}\vec{v}_{k+1} = \vec{0}$.) A consequence is that any linear operator with n distinct eigenvalues is diagonalizable.

(b) Let λ be an eigenvalue of a linear operator or matrix with characteristic polynomial $f(t)$. The **algebraic multiplicity** of λ is the largest postive integer k for which $(t - \lambda)^k$ is a factor of $f(t)$.

(c) Let λ be an eigenvalue of T. The subspace $E_\lambda = \{\vec{v} \in V \mid T(\vec{v}) = \lambda\vec{v}\} = \mathcal{N}([T]_\beta - \lambda I_n)$ is called the **eigenspace** of T associated with λ. This subspace consists of the zero vector and the eigenvectors of T corresponding to λ.

(d) Let λ be an eigenvalue of T with algebraic multiplicity m. Then $1 \le \dim(E_\lambda) \le m$. $\dim(E_\lambda)$ is called the **geometric multiplicity** of λ.

(e) Let T be a linear operator on V whose characteristic polynomial *splits*, i.e., factorizes into linear factors over F (equivalently, the algebraic multiplicites sum to $\dim V$), and $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then

i. T is diagonalizable iff the algebraic multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for all i.

ii. If T is diagonalizable and β_i is an ordered basis for E_{λ_i} for each i, then $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ is an ordered basis for V consisting of eigenvectors of T.

(f) Note that every polynomial over $\mathbb{C}$ splits, so every transformation over $\mathbb{C}$ has an eigenvalue.

(g) A more succinct condition for diagonalizability is as follows: a linear operator T on a finite-dimensional vector space V is diagonalizable iff V is the direct sum of the eigensapces of T.

(h) It follows that if $A \in M_{n\times n}(F)$ is a diagonalizable matrix, then $A = QDQ^{-1}$, where Q has as its columns the vectors in a basis of eigenvectors of A, and D has as its ith diagonal entry the eigenvalue of A corresponding to the ith column of Q.

$$A = \begin{bmatrix} | & & | \\ \vec{v_1} & \cdots & \vec{v_n} \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v_1} & \cdots & \vec{v_n} \\ | & & | \end{bmatrix}^{-1}$$

(i) Two linear operators T and U on V are **simultaneously diagonalizable** if there exists an ordered basis β for V such that $[T]_\beta$ and $[U]_\beta$ are diagonal matrices. Let T and U be diagonalizable. Then T and U are simultaneously diagonalizable iff T and U commute (i.e., $TU = UT$).

(j) Diagonalization can be used to efficiently compute powers of matrices. For example, consider the Fibonacci series. Let $\vec{v}_n = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}$, where F_n is the nth Fibonacci number (e.g., $\vec{v}_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$). Observe $\vec{v}_n = A^n \vec{v}_0$, where $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Diagonalizing A yields

$$A^n = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}$$

where $\lambda_1 = \frac{1-\sqrt{5}}{2}$ and $\lambda_2 = \frac{1+\sqrt{5}}{2}$. Simplifying and multiplying by $\vec{v}_0$ gives

$$\vec{v}_n = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_2^n - \lambda_1^n \\ \lambda_2^{n+1} - \lambda_1^{n+1} \end{bmatrix}$$

Hence a direct formula for F_n is

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right]$$

(k) We consider another application of diagonalization: solving systems of differential equations. Consider the system

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

where for each i, $x_i = x_i(t)$ is a differentiable real-valued function of t. Let

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}, \quad x' = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t) \end{bmatrix}$$

Then $Ax = x'$. A can be diagonalized into QDQ^{-1}, where

$$Q = \begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & -2 \\ 1 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Hence $x' = QDQ^{-1}$, or equivalently, $Q^{-1}x' = DQ^{-1}x$. Now let $y = Q^{-1}x$. Then $y' = Q^{-1}x' = Dy$. From this we get

$$\begin{aligned} y_1' &= 2y_1 \\ y_2' &= 2y_2 \\ y_3' &= 4y_3 \end{aligned}$$

The general solution to $x' = \lambda x$ is $x = ce^{\lambda t}$. So we have $y_1 = c_1 e^{2t}, y_2 = c_2 e^{2t}, y_3 = c_3 e^{4t}$. Finally, $x = Qy$, so we get

$$x(t) = e^{2t} \left(c_1 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right) + e^{4t} \left(c_3 \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} \right)$$

More generally, if A is the coefficient matrix of a system of differential equations, and A is diagonalizable with k distinct eigenvalues, the general solution is $x(t) = e^{\lambda_1 t}\vec{z_1} + e^{\lambda_2 t}\vec{z_2} + \ldots + e^{\lambda_k t}\vec{z_k}$ where $\vec{z_i} \in E_{\lambda_i}$.

14. **Invariant Subspaces**

(a) Throughout this section, let T be a linear operator on a vector space V. A subspace W of V is called a **T-invariant subspace** of V if $T(W) \subseteq W$, that is, if $T(\vec{v}) \in W$ for all $\vec{v} \in W$.

(b) Examples of T-invariant subspaces include $\{\vec{0}\}$, V, Im T, Ker T, and E_λ for any eigenvalue λ of T.

(c) If W and Z are T-invariant subspaces of V, then so is $W + Z$ and $W \cap Z$.

(d) For any polynomial $g(x)$, if W is T-invariant, W is also $g(T)$-invariant.

(e) Let $\vec{x}$ be a nonzero vector in V. The subspace

$$W = \text{Span}(\{\vec{x}, T(\vec{x}), T^2(\vec{x}), \ldots\})$$

is called the **T-cyclic subspace of V generated by $\vec{x}$**. W is the smallest T-invariant subspace of V containing $\vec{x}$.

(f) Let W be a T-invariant subspace of V. The *restriction*, denoted T_W, of T to W is a linear operator on W.

(g) Theorem: Let W be a T-invariant subspace of V. Then the characteristic polynomial of T_W divides the characteristic polynomial of T. *Proof:* Choose an ordered basis γ for W, and extend it to an ordered basis β of V. Let $B_1 = [T_W]_\gamma$. Then $[T]_\beta$ is of the form $\begin{bmatrix} B_1 & B_2 \\ O & B_3 \end{bmatrix}$, and $\text{char}_T(\lambda) = \text{char}_{T_W}(\lambda)\det(B_3 - \lambda I)$. Thus $\text{char}_{T_W}(\lambda)$ divides $\text{char}_T(\lambda)$.

(h) Theorem: Let W denote the T-cyclic subspace of V generated by a nonzero vector $\vec{v} \in V$, and $k = \dim(W)$. Then

i. $\{\vec{v}, T(\vec{v}), T^2(\vec{v}), \ldots, T^{k-1}(\vec{v})\}$ is a basis for W.

ii. If $a_0\vec{v} + a_1 T(\vec{v}) + \ldots + a_{k-1}T^{k-1}(\vec{v}) + T^k(\vec{v}) = \vec{0}$, then the characteristic polynomial of T_W is $f(t) = (-1)^k(a_0 + a_1 t + \ldots + a_{k-1}t^{k-1} + t^k)$.

(i) **Caley-Hamilton Theorem:** Let $f(t)$ be the characteristic polynomial of T. Then $f(T) = T_0$, the zero transformation. That is, T satisfies its characteristic equation. An analagous result holds for matrices.

(j) An implication of this theorem is that the set $\{I, T, T^2, \ldots, T^k\}$ is linearly dependent if $k \geq n$, since there is a nontrivial linear relation between $I, T, T^2, \ldots, T^n$ (the first coefficient of the characteristic polynomial is $(-1)^n \neq 0$). An example of an operator T that remains linearly independent up to T^{n-1} is one whose matrix has distinct entries on the diagonal.

15. **The Jordan Canonical Form**

(a) Now we consider matrix representations for non-diagonalizable operators, called *canonical forms*. The Jordan canonical form only requires that the characteristic polynomial splits, which recall always occurs if the underlying field is algebraically closed (e.g., the field of complex numbers).

(b) Overview: If a matrix is not diagonalizable, then at least one eigenspace is too "small"; so we extend the concept of an eigenspace to a *generalized* eigenspace. From these generalized eigenspaces, we select ordered bases whose union is an ordered basis β of V such that

$$[T]_\beta = \begin{bmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_k \end{bmatrix}$$

where O is the zero matrix and each A_i is a square matrix of the form

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$$

for some eigenvalue λ of T. Such a matrix A_i is called a **Jordan block** corresponding to λ. The matrix $[T]_\beta$ is called a **Jordan canonical form** of T, and β is a **Jordan canonical basis** for T. Notice $[T]_\beta$ is diagonal iff each A_i is of the form $\begin{bmatrix}\lambda\end{bmatrix}$.

(c) Every linear operator whose characteristic polynomial splits has a Jordan canonical form that is unique up to the order of the Jordan blocks.

(d) Let T be a linear operator on V. A nonzero vector $\vec{v} \in V$ is called a **generalized eigenvector** of T corresponding to eigenvalue λ if $(T - \lambda I)^p(\vec{v}) = \vec{0}$ for some sufficiently large positive integer p. Note that eigenvectors satisfy this condition for $p = 1$.

(e) The **generalized eigenspace of** T corresponding to λ, denoted K_λ, is defined by

$$K_\lambda = \{\vec{x} \in V \mid (T - \lambda I)^p(\vec{x}) = \vec{0} \text{ for some } p\}$$

This subspace consists of the zero vector and the generalized eigenvectors of T corresponding to λ. Alternatively, $K_\lambda = \mathcal{N}((T - \lambda I)^m)$ where m is the algebraic multiplicity of λ.

(f) K_λ is a T-invariant subspace of V containing E_λ. *Proof:* To see K_λ is a subspace, consider $\vec{x}, \vec{y} \in K_\lambda$. Then there exist p and q such that $(T - \lambda I)^p(\vec{x}) = (T - \lambda I)^q(\vec{y}) = \vec{0}$. Therefore

$$\begin{aligned}(T - \lambda I)^{p+q}(a\vec{x} + b\vec{y}) &= a(T - \lambda I)^q(\vec{0}) + b(T - \lambda I)^p(\vec{0}) \\ &= \vec{0}\end{aligned}$$

Hence $a\vec{x} + b\vec{y} \in K_\lambda$. To see K_λ is T-invariant, consider $\vec{x} \in K_\lambda$. Then there exists p such that $(T - \lambda I)^p(\vec{x}) = \vec{0}$. Then

$$(T - \lambda I)^p T(\vec{x}) = T(T - \lambda I)^p(\vec{x}) = T(\vec{0}) = \vec{0}$$

so that $T(\vec{x}) \in K_\lambda$. Here, we used the fact that T commutes with $(T - \lambda I)^m$ for any linear operator T and $m \geq 0$.

(g) The restriction of T onto K_λ has only one eigenvalue λ. *Proof:* Let μ be another eigenvalue of T_{K_λ}. Then $(T - \mu I)\vec{x} = \vec{0}$ for some eigenvector $\vec{x} \in K_\lambda$. Let p be the smallest integer for which $(T - \lambda I)^p(\vec{x}) = \vec{0}$, and let $\vec{y} = (T - \lambda I)^{p-1}(\vec{x})$. Hence $y \in E_\lambda$. Furthermore, $(T - \mu I)(\vec{y}) = (T - \mu I)(T - \lambda I)^{p-1}(\vec{x}) = (T - \lambda I)^{p-1}(T - \mu I)(\vec{x}) = \vec{0}$ so that $\vec{y} \in E_\mu$. But $E_\lambda \cap E_\mu = \{\vec{0}\}$, and thus $\vec{y} = \vec{0}$ and $\vec{x} = \vec{0}$, contradicting the assumption $\vec{x}$ was an eigenvector corresponding to μ.

(h) By a similar argument, we see if $\alpha \neq \beta$, then $K_\alpha \cap K_\beta = \{\vec{0}\}$. Therefore V can be written as a direct sum of the generalized eigenspaces of T, each one having dimension equal to the algebraic multiplicty of the corresponding eigenvalue.

(i) Let $\vec{x}$ be a generalized eigenvector of T corresponding to eigenvalue λ. Let p be the smallest postive integer for which $(T - \lambda I)^p(\vec{x}) = \vec{0}$. The length-$p$ cycle given by the ordered set

$$\{(T - \lambda I)^{p-1}(\vec{x}), \ldots, (T - \lambda I)(\vec{x}), \vec{x}\}$$

is called a *cycle of generalized eigenvectors* of T corresponding to λ. $(T - \lambda I)^{p-1}(\vec{x})$ is called the *initial vector* and $\vec{x}$ is called the *end vector*. Notice that the initial vector is the only eigenvector in the cycle, and that the cycle is linearly independent.

(j) Suppose β is a basis for V such that β is a disjoint union of cycles of generalized eigenvectors of T. Then

i. For each cycle γ of generalized eigenvectors in β, $W = \text{Span}(\gamma)$ is T-invariant, and $[T_W]_\gamma$ is a Jordan block.

ii. β is a Jordan canonical basis for V.

(k) The "dot" diagram of the JCF of a matrix is a table of dots that represent the cycles of generalized eigenvectors for each eigenvalue. For each eigenvalue, the number of columns is the number of Jordan blocks, and the total number of dots is the corresponding algebraic multiplicity.

(l) To Jordanalize a matrix A, it suffices to find a basis for each generalized eigenspace consisting of cycles of generalized eigenvectors. For simple cases, find a basis for each eigenspace, then, guided by the dot diagram, continually solve $(T - \lambda I)\vec{v}_{i+1} = \vec{v}_i$ to extend cycles as necessary. Finally place the bases together to form a Jordan canonical basis β; the matrix Q whose columns are the vectors of β will be such that $A = QJQ^{-1}$.

(m) This top-down method will not always work, as we might not be choosing eigenvectors at the top of a chain. A more robust method is as follows: for each eigenvalue λ_i, set $U = T - \lambda_i I$. Compute $d_j = \dim \operatorname{Ker}(U)^j$ for $1 \le j \le l$, where l is the stabilizing constant. For a given eigenvalue, d_j counts the number of dots in the first j rows. Construct a basis starting from the bottom row; after taking the images of the basis vectors on row $j+1$ as the leftmost basis vectors in row j, we need $(d_j - d_{j-1}) - (d_{j+1} - d_j) = 2d_j - d_{j-1} - d_{j+1}$ more basis vectors in $\operatorname{Ker}(U^j)$ to complete the basis for row j. This method shows that the JCF exists and is unique.

(n) The Caley-Hamilton theorem follows more easily using the Jordan canonical form. If $A = QJQ^{-1}$ and $f(t) = \operatorname{char}_A(t)$, then it can be shown that $f(A) = Qf(J)Q^{-1}$. Hence it suffices to show that $f(J) = O$, the zero matrix. Observe that $f(J) = \pm(J - \lambda_1 I)^{m_1}(J - \lambda_2 I)^{m_2} \cdots (J - \lambda_k I)^{m_k}$, i.e., the product of the characteristic polynomials of the restrictions of T to the generalized eigenspaces of T. Since $(J - \lambda_j I)^{m_j}(\vec{v}) = \vec{0}$ for $\vec{v} \in K_{\lambda_j}$, $f(J)$ is the zero operator on all K_{λ_j}. Since V is the direct sum of the subspaces K_{λ_j} for all j, it follows that $f(J) = O$.

16. **Inner Products**

(a) If $z = a + bi \in \mathbb{C}$, we define the complex conjugate $\overline{z} = a - bi$. The following properties hold:

- $z + \overline{z} = 2a = 2\Re(z)$ (real part of z)
- $z - \overline{z} = 2bi = 2i\Im(z)$ (imaginary part of z)
- $z \cdot \overline{z} = a^2 + b^2$
- $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$
- $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$
- $|z| = \sqrt{a^2 + b^2}$

(b) Let V be a vector space over F ($\mathbb{R}$ or $\mathbb{C}$). An **inner product** on V, denoted $\langle \vec{x}, \vec{y} \rangle$, is a function that assigns a scalar to every ordered pair of vectors $\vec{x}, \vec{y} \in V$ such that the following hold:

i. $\langle \vec{x} + \vec{z}, \vec{y} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{z}, \vec{y} \rangle$
ii. $\langle c\vec{x}, \vec{y} \rangle = c\langle \vec{x}, \vec{y} \rangle$
iii. $\langle \vec{x}, \vec{x} \rangle > 0$ if $\vec{x} \neq \vec{0}$
iv. $\overline{\langle \vec{x}, \vec{y} \rangle} = \langle \vec{y}, \vec{x} \rangle$

(c) Note that the last requirement reduces to $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$ if $F = \mathbb{R}$.

(d) For $\vec{x} = (a_1, \ldots, a_n)$, $\vec{y} = (b_1, \ldots, b_n) \in F^n$, define

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} a_i \overline{b_i}$$

This is called the **standard inner product** on F^n. For $F = \mathbb{R}$, the conjugation is not needed and this inner product is called the *dot product*, usually denoted $\vec{x} \cdot \vec{y}$.

(e) Example: Let V be the vector space of real-valued continuous functions on the interval $[0, 1]$. For $f, g \in V$, $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$ is an inner product.

(f) Let $A \in M_{m \times n}(F)$. Define the **conjugate transpose** or **adjoint** of A to be the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$ for all i, j. That is, $A^* = \overline{A^T}$.

(g) For $V = M_{n \times n}(F)$, the inner product defined by $\langle A, B \rangle = \operatorname{tr}(B^* A)$ for $A, B \in V$ is called the **Frobenius inner product**.

(h) A vector space V over F with a specific inner product is called an **inner product space**. Note that two distinct inner products on a given vector space yield two distinct inner product spaces. In addition to the properties required of an inner product, the following are true for any inner product space:

i. $\langle \vec{x}, \vec{y} + \vec{z} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{x}, \vec{z} \rangle$
ii. $\langle \vec{x}, c\vec{y} \rangle = \overline{c}\langle \vec{x}, \vec{y} \rangle$
iii. $\langle \vec{x}, \vec{0} \rangle = \langle \vec{0}, \vec{x} \rangle = \vec{0}$
iv. $\langle \vec{x}, \vec{x} \rangle = \vec{0}$ iff $\vec{x} = \vec{0}$
v. If $\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}, \vec{z} \rangle$ for all $\vec{x} \in V$, then $\vec{y} = \vec{z}$
Proof: Let β be a basis of V. First we show if $\langle \vec{x}, \vec{z} \rangle = 0$ for all $\vec{z} \in \beta$, then $\vec{x} = \vec{0}$. Write $\vec{x} = \sum_{i=1}^{n} a_i \vec{z}_i$. Then $\langle \vec{x}, \vec{x} \rangle = \langle \vec{x}, \sum_{i=1}^{n} a_i \vec{z}_i \rangle = \sum_{i=1}^{n} \overline{a_i} \langle \vec{x}, \vec{z}_i \rangle = 0$, which is true iff $\vec{x} = \vec{0}$. Substituting $\vec{x} - \vec{y}$ for $\vec{x}$ yields the result.

(i) Let V be an inner product space. For $\vec{x} \in V$, define the **norm** or **length** of $\vec{x}$ by $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$. In $\mathbb{R}^n$, this is the *magnitude* of a vector. For all $\vec{x}, \vec{y} \in V$ and $c \in F$, the following hold:

i. $\|c\vec{x}\| = |c| \cdot \|\vec{x}\|$
ii. $\|\vec{x}\| \ge 0$ with equality iff $\vec{x} = \vec{0}$

(j) All inner product spaces induce a norm, but we can also define a *normed space* as a vector space endowed with a norm satisfying the properties above. Some specific norms on $\mathbb{R}^n$ include

$$\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

$$\|\vec{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

Observe that the 1- and 2-norms are special cases of the p-norm, and the ∞-norm is the limit of the p-norm as p tend to infinity.

(k) Any norm induces a distance metric on V, given by $\|\vec{x} - \vec{y}\|$.

(l) *Cauchy-Schwarz Inequality*: $|\langle \vec{x}, \vec{y}\rangle| \leq \|\vec{x}\| \cdot \|\vec{y}\|$ (equality iff $\vec{x}$ and $\vec{y}$ are linearly dependent)
Proof: Assume $\vec{y} \neq \vec{0}$, since result is immediate otherwise. For any $c \in F$, we have

$$\begin{aligned} 0 \leq \|\vec{x} - c\vec{y}\|^2 &= \langle \vec{x} - c\vec{y}, \vec{x} - c\vec{y}\rangle \\ &= \langle \vec{x}, \vec{x} - c\vec{y}\rangle - c\langle \vec{y}, \vec{x} - c\vec{y}\rangle \\ &= \langle \vec{x}, \vec{x}\rangle - \bar{c}\langle \vec{x}, \vec{y}\rangle - c\langle \vec{y}, \vec{x}\rangle + c\bar{c}\langle \vec{y}, \vec{y}\rangle \end{aligned}$$

Taking $c = \frac{\langle \vec{x}, \vec{y}\rangle}{\langle \vec{y}, \vec{y}\rangle}$, the inequality becomes

$$0 \leq \langle \vec{x}, \vec{x}\rangle - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\langle \vec{y}, \vec{y}\rangle} = \|\vec{x}\|^2 - \frac{|\langle \vec{x}, \vec{y}\rangle|^2}{\|\vec{y}\|^2}$$

from which the result follows.

(m) *Triangle Inequality*: $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (equality iff $\vec{x}$ and $\vec{y}$ are positive scalar multiples)
Proof:

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y}\rangle \\ &= \langle \vec{x}, \vec{x}\rangle + \langle \vec{y}, \vec{x}\rangle + \langle \vec{x}, \vec{y}\rangle + \langle \vec{y}, \vec{y}\rangle \\ &= \|\vec{x}\|^2 + 2\Re\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2 \\ &\leq \|\vec{x}\|^2 + 2|\langle \vec{x}, \vec{y}\rangle| + \|\vec{y}\|^2 \\ &\leq \|\vec{x}\|^2 + 2\|\vec{x}\| \cdot \|\vec{y}\| + \|\vec{y}\|^2 \\ &= (\|\vec{x}\| + \|\vec{y}\|)^2 \end{aligned}$$

(n) Let V be an inner product space. Two vectors $\vec{x}, \vec{y} \in V$ are **orthogonal** if $\langle \vec{x}, \vec{y}\rangle = 0$. A vector $\vec{x} \in V$ is a **unit vector** if $\|\vec{x}\| = 1$. A subset S of V is **orthonormal** if any pair of distinct vectors in S are orthogonal and consists entirely of unit vectors.

(o) Note that for any nonzero vector $\vec{x}$, $(1/\|\vec{x}\|)\vec{x}$ is a unit vector; this process is called **normalizing**.

17. **Orthogonal Complements**

(a) Let V be an inner product space and $S = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthogonal basis of V. For $\vec{y} \in V$, we have

$$\vec{y} = \sum_{i=1}^{n} \frac{\langle \vec{y}, \vec{v}_i\rangle}{\|\vec{v}_i\|^2}\vec{v}_i$$

Proof: Write $\vec{y} = \sum_{i=1}^{n} a_i\vec{v}_i$. For $1 \leq j \leq n$, we have

$$\begin{aligned} \langle \vec{y}, \vec{v}_j\rangle = \left\langle \sum_{i=1}^{n} a_i\vec{v}_i, \vec{v}_j \right\rangle &= \sum_{i=1}^{n} a_i\langle \vec{v}_i, \vec{v}_j\rangle \\ &= a_j\langle \vec{v}_j, \vec{v}_j\rangle = a_j\|\vec{v}_j\|^2 \end{aligned}$$

So $a_j = \frac{\langle \vec{y}, \vec{v}_j\rangle}{\|\vec{v}_j\|^2}$. In the case S is an orthonormal basis, note that we have $\vec{y} = \sum_{i=1}^{n}\langle \vec{y}, \vec{v}_i\rangle\vec{v}_i$.

(b) The **Gram-Schmidt** process converts a linearly independent set into an orthogonal set. Let V be an inner product space and $S = \{\vec{w}_1, \ldots, \vec{w}_n\}$ be a linearly independent subset of V. Define $S' = \{\vec{v}_1, \ldots, \vec{v}_n\}$ where $\vec{v}_1 = \vec{w}_1$ and

$$\vec{v}_k = \vec{w}_k - \sum_{j=1}^{k-1} \frac{\langle \vec{w}_k, \vec{v}_j\rangle}{\|\vec{v}_j\|^2}\vec{v}_j \qquad \text{for } 2 \leq k \leq n$$

Then S' is an orthogonal set such that $\mathrm{Span}(S') = \mathrm{Span}(S)$.

(c) Let $V = \mathbb{P}_n(\mathbb{R})$ with inner product $\langle f(x), g(x)\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Applying the Gram-Schmidt process on the standard basis $\{1, x, x^2, \ldots\}$ produces an orthogonal basis whose elements are called the *Legendre polynomials*.

(d) A consequence of Gram-Schmidt is that every nonzero finite-dimensional inner product space has an orthonormal basis.

(e) Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{\vec{v}_1, \ldots, \vec{v}_n\}$. Let T be a linear operator on V, and $A = [T]_\beta$. Then for any i and j, $A_{ij} = \langle T(\vec{v}_j), \vec{v}_i\rangle$.

(f) Let β be an orthonormal subset of an inner product space V, and let $\vec{x} \in V$. Define the **Fourier** coefficients of $\vec{x}$ relative to β to be the scalars $\langle \vec{x}, \vec{y}\rangle$, where $\vec{y} \in \beta$. Consider the vector space V of functions over $\mathbb{C}$ on the interval $[0, 2\pi]$ with inner product $\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt$. The set $\{1, \cos nt, \sin nt \mid n \in \mathbb{N}\}$ is orthogonal on V, and is the basis from which the "classical" Fourier coefficients arise.

(g) Let S be a nonempty subset of an inner product space V. The **orthogonal complement** $S^\perp$ of S to be the set of all vectors in V that are orthogonal to every vector in S; i.e., $S^\perp = \{\vec{x} \in V \mid \langle \vec{x}, \vec{y}\rangle = 0 \quad \forall \vec{y} \in S\}$.

(h) If S is a subset of V, then $S^\perp$ is a subspace of V. For any subspace W of V, $W \cap W^\perp = \{\vec{0}\}$, $(W^\perp)^\perp = W$, $\{\vec{0}\}^\perp = V$, and $V = W \oplus W^\perp$.

(i) For any two subspaces W_1 and W_2, $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$ and $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$.

(j) Let W be a finite-dimensional subspace of an inner product space V, and let $\vec{y} \in V$. Then there exist unique vectors $\vec{u} \in W$ and $\vec{z} \in W^\perp$ such that $\vec{y} = \vec{u} + \vec{z}$. Furthermore, if $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is an orthonormal basis for W, then

$$\vec{u} = \sum_{i=1}^{k}\langle \vec{y}, \vec{v}_i\rangle\vec{v}_i$$

The vector $\vec{u}$ is the unique vector in W that is closest to $\vec{y}$; that is, $\|\vec{y} - \vec{x}\| \geq \|\vec{y} - \vec{u}\|$ for $\vec{x} \in W$, with equality iff $\vec{x} = \vec{u}$. It is called the **orthogonal projection** of $\vec{y}$ onto W.

(k) Let $A \in M_{n\times n}(\mathbb{C})$. $AA^* = I_n$ iff the rows of A are orthonormal. *Proof:* Recall AB_{ij} is the dot product of row i of A and column j of B. So the ij^{th} entry of AA^* is the dot product of row i of A and column j of A^*, which is the conjugate of row j of A. Denoting the rows of A as $a_1, \ldots, a_n$, we have $AA^* = I_n$ iff $\langle a_i, \overline{a_j}\rangle$ is 0 when $i \neq j$ and 1 when $i = j$.

Furthermore, since the rows are orthonormal, A is invertible and $A^{-1} = A^*$. Thus we also have $A^*A = I_n$, and by the same argument as above, the columns of A are orthonormal.

(l) For a matrix $A \in M_{m\times n}(F)$, $\mathcal{N}(A)^\perp = R(A^T)$ and $\mathcal{N}(A^T)^\perp = R(A)$ where $\mathcal{N}(A)$ and $R(A)$ denote the null space and range of A, respectively. This is seen in that the equality $A\vec{x} = \vec{0}$ means $\vec{x}$ is orthogonal to the rows of A.

18. **The Adjoint of a Linear Operator**

(a) Let $g : V \mapsto F$ be a linear functional on a finite-dimensional inner product space V over F. Then there exists a unique vector $\vec{y} \in V$ such that $g(\vec{x}) = \langle \vec{x}, \vec{y}\rangle$ for all $\vec{x} \in V$. *Proof:* If $\beta = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal basis for V, take $\vec{y} = \sum_{i=1}^n \overline{g(\vec{v}_i)}\vec{v}_i$.

(b) Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Then there exists a unique linear operator $T^* : V \mapsto V$ such that $\langle T(\vec{x}), \vec{y}\rangle = \langle \vec{x}, T^*(\vec{y})\rangle$ for all $\vec{x}, \vec{y} \in V$.

(c) The operator T^* is called the **adjoint** of T. Note that we also have $\langle \vec{x}, T(\vec{y})\rangle = \langle T^*(\vec{x}), \vec{y}\rangle$ for all $\vec{x}, \vec{y} \in V$.

(d) Let V be a finite-dimensional inner product space, and β an *orthonormal* basis for V. If T is a linear operator on V, then

$$[T^*]_\beta = [T]_\beta^*$$

(e) We have the following properties (with similar results holding for matrices):

i. $(T + U)^* = T^* + U^*$
ii. $(cT)^* = \overline{c}T^*$ for any $c \in F$
iii. $(TU)^* = U^*T^*$
iv. $T^{**} = T$
v. If T is invertible, $(T^*)^{-1} = (T^{-1})^*$.

(f) If $A \in M_{n\times n}(F)$, then $\det(A^*) = \overline{\det(A)}$.

(g) Let T be a linear operator on an inner product space V. Then $\|T(\vec{x})\| = \|\vec{x}\|$ for all $\vec{x} \in V$ iff $\langle T(\vec{x}), T(\vec{y})\rangle = \langle \vec{x}, \vec{y}\rangle$ for all $\vec{x}, \vec{y} \in V$. *Proof:* The backwards direction is trivial. For the forward direction, suppose $\|T(\vec{x})\| = \|\vec{x}\|$; expand $\|T(\vec{x}+\vec{y})\|$ by linearity. We arrive at $\Re(\langle T(\vec{x}), T(\vec{y})\rangle) = \Re(\langle \vec{x}, \vec{y}\rangle)$. Applying the same argument to $\vec{x} + i\vec{y}$ shows that the imaginary parts are also equal, and the claim follows. A consequence is that linear operators that preserve lengths (e.g. reflections and rotations in the plane) also preserve angles between vectors (we can define the angle θ between $\vec{x}$ and $\vec{y}$ as $\cos\theta = \frac{\langle \vec{x}, \vec{y},\rangle}{\|\vec{x}\|\|\vec{y}\|}$).

(h) Let $A \in M_{m\times n}(F)$. Then $\text{rank}(A^*A) = \text{rank}(A)$. As a result, if A is an $m \times n$ matrix such that $\text{rank}(A) = n$, then $A^*A \in M_{n\times n}(F)$ is invertible.

(i) **Least Squares Approximation:** Let $A \in M_{m\times n}(F)$ and $\vec{y} \in F^m$. Then there exists $\vec{x}_0 \in F^n$ such that $(A^*A)\vec{x}_0 = A^*\vec{y}$ and $\|A\vec{x}_0 - \vec{y}\| \leq \|A\vec{x} - \vec{y}\|$ for all $\vec{x} \in F^n$.
Proof: From the discussion on orthogonal projections, there exists a unique vector in $R(A)$ that is closest to $\vec{y}$. Call this vector $A\vec{x}_0$, where $\vec{x}_0 \in F^n$. To compute $\vec{x}_0$, note that $A\vec{x}_0 - \vec{y} \in R(A)^\perp$, so $\langle A\vec{x}, A\vec{x}_0 - \vec{y}\rangle = 0$ for all $\vec{x} \in F^n$. This is equivalent to $\langle \vec{x}, A^*(A\vec{x}_0 - \vec{y})\rangle = 0$ for all $\vec{x} \in F^n$; that is, $A^*(A\vec{x}_0 - \vec{y}) = 0$. Hence we need only find a solution $\vec{x}_0$ to $A^*A\vec{x} = A^*\vec{y}$. If $\text{rank}(A) = n$, then $\vec{x}_0$ is unique.

(j) An important application of least squares is curve fitting. Suppose we have data points $(x_1, y_1), \ldots, (x_m, y_m)$. If we wish to fit the polynomial $y = ax^2 + bx + c$ to the data, for instance, the appropriate model is

$$A = \begin{bmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_m^2 & x_m & 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Applying least squares to the system $A\vec{x} = \vec{b}$ yields the coefficients a, b, c that minimize error.

(k) A solution $\vec{s}$ to $A\vec{x} = \vec{b}$ is called a *minimal solution* if $\|\vec{s}\| \leq \|\vec{u}\|$ for all other solutions $\vec{u}$ (i.e., the solution of minimal norm). Every consistent system of linear equations has a unique minimal solution $\vec{s}$. Let $\vec{u}$ be a solution to $AA^*\vec{x} = \vec{b}$. Then $\vec{s} = A^*\vec{u}$.

19. **Normal and Self-Adjoint Operators**

(a) Let V be an inner product space, and let T be a linear operator on V. T is called **normal** if it commutes with its adjoint, that is $TT^* = T^*T$. Similarly, a $n \times n$ matrix A is called normal if $AA^* = A^*A$.

(b) Let T be a normal operator on V. Then

i. $\|T(\vec{x})\| = \|T^*(\vec{x})\|$ for all $\vec{x} \in V$. *Proof:* For any $\vec{x} \in V$, $\|T(\vec{x})\|^2 = \langle T(\vec{x}), T(\vec{x})\rangle = \langle T^*T(\vec{x}), \vec{x}\rangle = \langle TT^*(\vec{x}), \vec{x}\rangle = \langle T^*(\vec{x}), T^*(\vec{x})\rangle = \|T^*(\vec{x})\|^2$
ii. $T - cI$ is normal $\forall c \in F$
iii. If $T(\vec{x}) = \lambda\vec{x}$, then $T^*(\vec{x}) = \overline{\lambda}\vec{x}$. *Proof:* Suppose that $T(\vec{x}) = \lambda\vec{x}$; then $(T - \lambda I)(\vec{x}) = \vec{0}$.

Thus $0 = \|(T - \lambda I)(\vec{x})\| = \|(T - \lambda I)^*(\vec{x})\| = \|(T^* - \overline{\lambda} I)(\vec{x})\| = \|T^*(\vec{x}) - \overline{\lambda}\vec{x}\|$. Hence $T^*(\vec{x}) = \overline{\lambda}\vec{x}$.

iv. If λ_1 and λ_2 are distinct eigenvalues of T with corrsponding eigenvectors $\vec{x}_1$ and $\vec{x}_2$, then $\vec{x}_1$ and $\vec{x}_2$ are orthogonal. *Proof:* $\lambda_1\langle\vec{x}_1, \vec{x}_2\rangle = \langle\lambda_1\vec{x}_1, \vec{x}_2\rangle = \langle T(\vec{x}_1), \vec{x}_2\rangle = \langle\vec{x}_1, T^*(\vec{x}_2)\rangle = \langle\vec{x}_1, \overline{\lambda_2}\vec{x}_2\rangle = \lambda_2\langle\vec{x}_1, \vec{x}_2\rangle$. Since $\lambda_1 \neq \lambda_2$, $\langle\vec{x}_1, \vec{x}_2\rangle = 0$.

(c) Example: For the rotation matrix A for $\mathbb{R}^2$, we have $AA^* = I = A^*A$, so A is normal. For a real skew-symmetric matrix A, we have $AA^T = -A^2 = A^TA$, so A is normal.

(d) Let T be a linear operator on a finite-dimensional complex inner product space V. Then T is normal iff there exists an orthonormal basis for V consisting of eigenvectors of T. Intuitively, if β is an orthonormal basis of eigenvectors, then we see $[T]_\beta$ is diagonal, and so is $[T^*]_\beta$. Diagonal matrices commute, so T and T^* commute.

(e) Normality is not sufficient to guarantee the existence of an orthonormal basis of eigenvectors for real inner product spaces. We replace normality by the stronger condition that $T = T^*$ in order to guarantee such a basis.

(f) Let T be a linear operator on an inner product space V. T is called **self-adjoint** if $T = T^*$. Similarly, a $n \times n$ matrix A is called self-adjoint or **Hermitian** if $A = A^*$. Note for real matrices, this reduces to the requirement that A is symmetric.

(g) Let T be a self-adjoint operator on a finite-dimensional inner product space V. Then every eigenvalue of T is real. *Proof:* Since a self-adjoint operator is also normal, $\lambda\vec{x} = T(\vec{x}) = T^*(\vec{x}) = \overline{\lambda}\vec{x}$ so $\lambda = \overline{\lambda}$, that is, λ is real. As a result, the characteristic polynomial of T splits over $\mathbb{R}$.

(h) Let T be a linear operator on a finite-dimensional real inner product space V. Then T is self-adjoint iff there exists an orthonormal basis for V consisting of eigenvectors of T.

(i) A simple corollary is that any Hermitian (over $\mathbb{C}$) or symmetric (over $\mathbb{R}$) matrix is diagonalizable in some orthonormal basis. Indeed, suppose a real symmetric matrix $A = QDQ^T$, where the columns of Q are orthonormal eigenvectors $\vec{u}_1, \ldots, \vec{u}_k$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$. Using column-row expansion, we can write

$$A = \lambda_1\vec{u}_1\vec{u}_1^T + \lambda_2\vec{u}_2\vec{u}_2^T + \ldots + \lambda_k\vec{u}_k\vec{u}_k^T$$

This representation of A is called a **spectral decomposition** of A.

20. **Unitary and Orthogonal Operators**

(a) Let T be a linear operator on a finite-dimensional inner product space V. If $\|T(\vec{x})\| = \|\vec{x}\|$ for all $\vec{x} \in V$, we call T a **unitary operator** (over $\mathbb{C}$) or an **orthogonal operator** (over $\mathbb{R}$).

(b) A square matrix A is called an **orthogonal matrix** if $A^TA = AA^T = I$ and **unitary** if $A^*A = AA^* = I$.

(c) A linear operator T is normal, self-adjoint, unitary, or orthogonal iff its matrix $[T]_\beta$ if normal, self-adjoint, unitary, or orthogonal, repsectively where β is an *orthonormal* basis.

(d) There are several equivalent definitions of unitary or orthogonal operators (or matrices). If T is a linear operator on an inner product space V, the following are equivalent:

i. $TT^* = T^*T = I$. Recall from earlier that this is equivalent to the rows (and columns) of the matrix being orthonormal.

ii. $\langle T(\vec{x}), T(\vec{y})\rangle = \langle\vec{x}, \vec{y}\rangle$ for all $\vec{x}, \vec{y} \in V$. This follows from (i) since $\langle\vec{x}, \vec{y}\rangle = \langle T^*T(\vec{x}), \vec{y}\rangle = \langle T(\vec{x}), T(\vec{y})\rangle$ for any $\vec{x}, \vec{y} \in V$.

iii. $\|T(\vec{x})\| = \|\vec{x}\|$ for all $\vec{x} \in V$. The equivalence to (ii) was shown earlier.

iv. If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V.

(e) It follows that every eigenvalue of a unitary or orthogonal operator has absolute value 1.

(f) Matrices A and B are **unitarily equivalent** over $\mathbb{C}$ (or orthogonally equivalent over $\mathbb{R}$) if $A = P^*BP$ for some unitary or orthogonal matrix P. Note that $P^* = P^{-1}$ so that $A \sim B$.

(g) For any complex normal (or real symmetric) matrix A, we have $A = QDQ^*$ where Q is unitary (or orthogonal). The converses are also true, so that we have

i. Let $A \in M_{n\times n}(\mathbb{C})$. Then A is normal iff A is unitarily equivalent to a diagonal matrix.

ii. Let $A \in M_{n\times n}(\mathbb{R})$. Then A is symmetric iff A is orthogonally equivalent to a diagonal matrix.

(h) Let V be a real inner product space. A function $f : V \mapsto V$ is called a **rigid motion** if $\|f(\vec{x}) - f(\vec{y})\| = \|\vec{x} - \vec{y}\|$ for all $\vec{x}, \vec{y} \in V$. In other words, it is a transformation that preserves distances.

(i) It is clear that any orthogonal operator or translation on a finite-dimensional real inner product space is a rigid motion. In fact, every rigid motion is a composite of these two rigd motions.

(j) Let $f : V \mapsto V$ be a rigid motion; then there exists a unique orthogonal operator T on V and a unique translation g on V such that $f = g \circ T$.

21. **Singular Value Decomposition**

(a) The singular value decomposition is a useful factorization of any $m \times n$ matrix in applied linear algebra. We will work with real matrices from now on.

(b) A symmetric matrix A is **positive definite**, denoted $A \succ 0$, if for all nonzero $\vec{x} \in \mathbb{R}^n$, $\vec{x}^T A\vec{x} > 0$. This is equivalent to having all positive eigenvalues.

(c) A symmetric matrix A is **positive semi-definite**, denoted $A \succeq 0$, if for all vectors $\vec{x}^T A\vec{x} \geq 0$. This is equivalent to having all nonnegative eigenvalues.

(d) **Negative definite** and **negative semi-definite matrices** are defined similarly, and a symmetric matrix is **indefinite** if it is neither positive semi-definite nor negative semi-definite.

(e) Let $A \in M_{m\times n}(\mathbb{R})$. Then $A^T A$ is symmetric and thus orthogonally diagonalizable. Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$, and $\lambda_1, \ldots, \lambda_n$ be the associated eigenvalues. For $1 \leq i \leq n$,

$$\|A\vec{v}_i\|^2 = (A\vec{v}_i)^T A\vec{v}_i = \vec{v}_i^T A^T A\vec{v}_i = \vec{v}_i^T(\lambda_i \vec{v}_i) = \lambda_i$$

The eigenvalues of $A^T A$ (sometimes called the Gram matrix) are nonnegative, so $A^T A$ is positive semi-definite. The **singular values** of A, denoted $\sigma_1, \ldots, \sigma_n$, are the square roots of the eigenvalues of $A^T A$ arranged in decreasing order.

(f) The number of nonzero singular values of A is equal to the rank of A.

(g) Let $A \in M_{m\times n}(\mathbb{R})$ with rank r and positive singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$. Let Σ be the $m \times n$ matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i \text{ if } i = j \leq r \\ 0 \text{ otherwise} \end{cases}$$

Then there exists an $m\times m$ matrix orthogonal matrix U and an $n \times n$ matrix orthogonal matrix V such that

$$A = U\Sigma V^T$$

This is called the **singular value decomposition** of A.

(h) The columns of U are called the left-singular vectors of A, and they are unit eigenvectors of AA^T. The columns of V are called the right-singular vectors of A, and they are unit eigenvectors of $A^T A$.

(i) Let $A \in M_{m\times n}(\mathbb{R})$ with rank r and SVD $A = U\Sigma V^T$. Let $\Sigma^\dagger$ be the $n \times m$ matrix defined by

$$\Sigma^\dagger_{ij} = \begin{cases} \frac{1}{\sigma_i} \text{ if } i = j \leq r \\ 0 \text{ otherwise} \end{cases}$$

Then $A^\dagger = V\Sigma^\dagger U^T$. $A^\dagger$ is called the **pseudoinverse** of A.

(j) Consider the system of linear equations $A\vec{x} = \vec{b}$, where A is an $m \times n$ matrix and $\vec{b} \in \mathbb{R}^m$. If $\vec{z} = A^\dagger \vec{b}$, then

i. if $A\vec{x} = \vec{b}$ is consistent, then $\vec{z}$ is the unique solution to the system having minimum norm.

ii. if $A\vec{x} = \vec{b}$ is inconsistent, then $\vec{z}$ is the unique best approximation to a solution having minimum norm.

22. **Quadratic Forms on $\mathbb{R}^n$**

(a) A quadratic form on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \mapsto \mathbb{R}$ of the form $Q(\vec{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j$.

(b) There is a one-to-one correspondence between quadratic forms and symmetric matrices; any quadratic form Q can be written $Q(\vec{x}) = \vec{x}^T A\vec{x}$, where A is an $n \times n$ symmetric matrix. Using the above notation, $A_{ij} = a_{ij}$.

(c) In practice, we can reconstruct Q from A as folows:

$$Q(\vec{x}) = \sum_{i=1}^n A_{ii}x_i^2 + \sum_{i<j} 2A_{ij}x_i x_j$$

(d) A simple example of a nonzero quadratic form is $Q(\vec{x}) = x_1^2 + x_2^2 + \ldots + x_n^2 = \|\vec{x}\|^2 = \vec{x}^T I_n \vec{x}$.

(e) Q is positive definite if $Q(\vec{x}) > 0$ for all $\vec{x} \neq \vec{0}$, which is equivalent to its matrix A having all positive eigenvalues. Other terms are defined analogously.

(f) Let A be an $n \times n$ matrix. Then A is positive definite iff $\det(A_k) > 0$ for $1 \leq k \leq n$, where here A_k denotes the $k \times k$ submatrix formed from the first k rows and first k columns of A.

(g) Diagonalizing the matrix of a quadratic form yields a simpler representation of the quadratic form. For example, consider the equation

$$3x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1x_2 + 2\sqrt{2}(x_1 + x_3) + 1 = 0$$

Let

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 3 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 2\sqrt{2} \\ 0 \\ 2\sqrt{2} \end{bmatrix}$$

So that $\vec{x}^T A\vec{x} + \vec{u} \cdot \vec{x} + 1 = 0$. An orthonormal basis of eigenvectors for A is $\frac{1}{\sqrt{2}}(1, 0, 1)$, $(0, 1, 0)$, $\frac{1}{\sqrt{2}}(1, 0, -1)$ for eigenvalues 2, 3, 4, respectively. In this basis, the equation becomes

$$2x_1^2 + 3x_2^2 + 4x_3^2 + 4x_1 + 1 = 0$$
$$\implies 2(x_1 + 1)^2 + 3x_2^2 + 4x_3^2 = 1$$

which is an ellipsoid centered at $(-1, 0, 0)$.