Linear Algebra

SF1672

Contents

1. Vector Spaces
   - 1.1. Definitions
   - 1.2. Subspaces
   - 1.3. Bases
   - 1.4. Dimension
2. Matrices
   - 2.1. Row Reduction
     - 2.1.1. Echelon Form
     - 2.1.2. LU Decomposition
     - 2.1.3. The Column Space
     - 2.1.4. Matrix Inverse
   - 2.2. Determinants
     - 2.2.1. Permutations
     - 2.2.2. Computation
     - 2.2.3. Linear Systems
     - 2.2.4. Adjugates
     - 2.2.5. Area and Volume
3. Linear Maps
   - 3.1. Definitions
   - 3.2. Properties
   - 3.3. In Vector Spaces
4. Sequences
   - 4.1. Linear
   - 4.2. Repetition of Real Analysis
     - 4.2.1. Definitions
     - 4.2.2. Bounded Sequences
     - 4.2.3. Cauchy Sequences
5. Orthogonality
   - 5.1. Inner Products
   - 5.2. Vectors
   - 5.3. Projection
     - 5.3.1. Intersections
     - 5.3.2. Orthogonal Basis
     - 5.3.3. Approximation
   - 5.4. Sets
   - 5.5. Matrices
6. Eigenvalues and Eigenvectors
   - 6.1. Of Matrices
   - 6.2. Real
     - 6.2.1. Diagonalization
     - 6.2.2. Maps
     - 6.2.3. The Real Spectral Theorem
   - 6.3. Quadratic Forms
     - 6.3.1. Definition
     - 6.3.2. Principal Axes
     - 6.3.3. Definiteness
   - 6.4. Complex
     - 6.4.1. Complex Calculations
     - 6.4.2. Hermitian Spaces
     - 6.4.3. Unitary Bases
     - 6.4.4. The Complex Spectral Theorem
7. Statistics
   - 7.1. Linear Regression
   - 7.2. Markov Chains

1. Vector Spaces

1.1. DEFINITIONS

Definition 1.1.1 (Group)

A *group* is a set G with a binary operation $*$ such that for all $a, b, c \in G$,

(i) (Closure) $a * b \in G$
(ii) (Associativity) $(a * b) * c = a * (b * c)$
(iii) (Identity) $\exists e \in G : e * a = a * e = a$
(iv) (Inverse) $\exists a^{-1} \in G : a * a^{-1} = a^{-1} * a = e$

Additionally, a group is said to be *abelian* if

(v) (Commutative) $a * b = b * a$

Note that the proof of (iii) and (iv) can be shortened if (v) is proven.

Definition 1.1.2 (Field)

A *field* is a set F with two binary operations $+$ and $\cdot$ such that

(i) $(F, +)$ is an abelian group with identity 0.
(ii) $(F \setminus \{0\}, \cdot)$ is an abelian group with identity 1.
(iii) Multiplication distributes over addition:
for all $a, b, c \in F$, $a \cdot (b + c) = a \cdot b + a \cdot c$.

Examples. $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{R}(x), \mathbb{Q}(i), \mathbb{F}_p$ if p is prime.

Definition 1.1.3 (Vector Space)

A vector space over a field F is a set V with two operations:

- Vector addition $+ : V \times V \to V$
- Scalar multiplication $\cdot : F \times V \to V$

These must satisfy

(i) (Closure under addition) $\mathbf{u} + \mathbf{v} \in V$
(ii) (Distributivity over vectors) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
(iii) (Associativity) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
(iv) (Identity) $\mathbf{u} + \mathbf{0} = \mathbf{u}$

(v) (Inverse) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
(vi) (Commutativity) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(vii) (Closure under scalar multiplication) $a\mathbf{u} \in V$
(viii) (Distributivity over scalars) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$
(ix) (Associativity) $a(b\mathbf{u}) = (ab)\mathbf{u}$
(x) (Identity) $1\mathbf{u} = \mathbf{u}$

Note that requiring closure is often omitted.

1.19 Definition ***vector space***

A ***vector space*** is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

commutativity
$u + v = v + u$ for all $u, v \in V$;

associativity
$(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbf{F}$;

additive identity
there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$;

additive inverse
for every $v \in V$, there exists $w \in V$ such that $v + w = 0$;

multiplicative identity
$1v = v$ for all $v \in V$;

distributive properties
$a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbf{F}$ and all $u, v \in V$.

1.2. SUBSPACES

Definition 1.2.1 (Subspace)

A subspace of a vector space V is a subset H that is

(i) nonempty (e.g., $\mathbf{0} \in H$),
(ii) closed under addition, and
(iii) closed under scalar multiplication.

Example. The set $\{\mathbf{0}\}$ (with $\mathbf{0} \in V$) is a subspace of every V.

⚠ Warning

$\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$. However, the set $\{(s, t, 0) : s, t \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$.

Definition 1.2.2 (Span)

The span of a set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is the set of all linear combinations of the vectors in S. The span of a set of vectors is the smallest subspace that contains them.

Theorem 1.2.1

If $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$, then $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

1.3. BASES

Definition 1.3.1 (Basis)

A basis for a vector space V is a set $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of linearly independent vectors that span V.

Example. The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$, and the set $\{1, t, t^2, \ldots, t^n\}$ is the standard basis for P_n.

Theorem 1.3.1 (Basis Criterion)

Let V be an n-dimensional vector space with $n \geq 1$. Any set of n vectors in V is automatically a basis if

(i) it is linearly independent or, equivalently,
(ii) it spans V.

Definition 1.3.2 (Coordinates)

Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for V. Each geometric vector $\mathbf{v} \in V$ can be expressed as a coordinate vector

$$[\mathbf{v}]_B = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_B,$$

the injective coordinate mapping $\mathbf{v} \to [\mathbf{v}]_B$,

$$P_B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$$

from B to the standard basis in V, with

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

Now let $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be another basis for V. Then there is a unique $n \times n$ matrix

$$P_{C \leftarrow B} = \left[[\mathbf{b}_1]_C \ \cdots \ [\mathbf{b}_n]_C\right]$$

$$[\mathbf{x}]_C = P_{C \leftarrow B}[\mathbf{x}]_B$$

and

$$(P_{C \leftarrow B})^{-1} = P_{B \leftarrow C}.$$

Problem. For a vector space in $\mathbb{R}^n$, let E be the standard basis and the bases B and C be given. Find the change-of-basis matrix from B to C.

Solution. The change-of-basis matrices give

$$P_{C \leftarrow B} = P_{C \leftarrow E} P_{E \leftarrow B} = (P_{E \leftarrow C})^{-1} P_{E \leftarrow B}.$$

Alternative solution. For simplicity assume that the dimension is $n = 2$. The bases of B in C-coordinates $\mathbf{v}_1$ and $\mathbf{v}_2$ satisfy

$$[\mathbf{c}_1 \ \mathbf{c}_2]_E [\mathbf{v}_1 \ \mathbf{v}_2]_C = [\mathbf{b}_1 \ \mathbf{b}_2]_E [I]_B,$$

thus being given by reducing

$$[\mathbf{c}_1 \ \mathbf{c}_2 \mid \mathbf{b}_1 \ \mathbf{b}_2] \sim [I \mid \mathbf{v}_1 \ \mathbf{v}_2].$$

1.4. DIMENSION

Definition 1.4.1 (Dimension)

The dimension of a vector space is the number of vectors in every basis. A vector space is either finite-dimensional och infinite-dimensional.

Theorem 1.4.1

With matrix addition and scalar multiplication, $\dim F^{m \times n} = mn$.

Definition 1.4.2 (Rank and Nullity)

The rank of a linear map (or matrix) is the dimension of its image (column space) and is also given by the number of pivot columns, and the nullity is the dimension of its kernel.

Theorem 1.4.2 (Rank Theorem)

For an $m \times n$ matrix A it holds that

$$\operatorname{rank} A + \operatorname{nullity} A = n.$$

2. Matrices

2.1. ROW REDUCTION

2.1.1. Echelon Form

Definition 2.1.1 (Echelon Form)

- Row Echelon Form (REF):
  Pivots move to the right as you go down, with zeros below each pivot.
- Reduced Row Echelon Form (RREF):
  REF plus each pivot is 1 and is the only nonzero entry in its column. Canonical, i.e., unique.

⚠ Warning

Some authors also require pivots to start with 1 for REF.

Definition 2.1.2 (Consistent)

A linear system $A\mathbf{x} = \mathbf{b}$ is consistent if it has at least one solution. That is, $\mathbf{b} \in \operatorname{col} A$.

Theorem 2.1.1 (Echelon Test)

After reducing $[A \mid \mathbf{b}]$ to REF or RREF, a system is inconsistent if and only if you obtain a row of the form

$$[0 \ 0 \ \cdots \ 0 \mid c], \quad c \neq 0.$$

2.1.2. LU Decomposition

Definition 2.1.3 (LU Decomposition)

A square matrix A admits an *LU decomposition* if it can be factored as

$$A = LU,$$

where L is lower triangular (often with ones on the diagonal) and U is upper triangular.

Theorem 2.1.2 (LU with Pivoting)

In general one uses partial pivoting:

$$PA = LU,$$

where P is a permutation matrix.

💡 Solving Systems

$A\mathbf{x} = \mathbf{b}$ solved by $PA = LU$:

(i) Compute $PA = LU$ once.
(ii) Solve $L\mathbf{y} = P\mathbf{b}$ (forward substitution).
(iii) Solve $U\mathbf{x} = \mathbf{y}$ (back substitution).

ⓘ Note

Solving systems through substitution after LU decomposition is $\mathcal{O}(n^2)$ while Gaussian elimination of the original matrix is $\mathcal{O}(n^3)$.

2.1.3. The Column Space

Definition 2.1.4 (Column Space)

$$\operatorname{col}(A) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \operatorname{im}(f_A),$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are the columns of A.

Theorem 2.1.3 (Pivot Basis)

The pivot columns of a matrix form a basis for its column space.

Theorem 2.1.4

The column space is affected by row reduction but not by column reduction, and vice versa for the row space.

2.1.4. Matrix Inverse

Definition 2.1.5 (Singular)

A matrix is singular if it is square and non-invertible.

Definition 2.1.6 (Elementary Matrix)

An elementary matrix is an $n \times n$ matrix obtained by applying one elementary row operation to the identity matrix I_n.

All elementary matrices are invertible.

Theorem 2.1.5 (Invertible)

Let A be an $n \times n$ matrix. Then it is equivalent that

(i) A is invertible
(ii) A^{T} is invertible
(iii) A has a left-hand inverse
(iv) A has a right-hand inverse

(v) A has n pivot positions
(vi) A is row equivalent to the $n \times n$ identity matrix
(vii) A can be expressed as a product of elementary matrices
(viii) $A\mathbf{x} = \mathbf{b}$ has at least (and thus exactly) one solution for each $\mathbf{b}$
(ix) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution
(x) 0 is *not* an eigenvalue of A

(xi) the columns of A are linearly independent, forming a basis for $\mathbb{R}^n$
(xii) $\operatorname{col} A = \mathbb{R}^n$
(xiii) $\ker A = \{\mathbf{0}\}$
(xiv) $\operatorname{rank} A = n$
(xv) $\operatorname{nullity} A = 0$
(xvi) $(\operatorname{col} A)^{\perp} = \{\mathbf{0}\}$
(xvii) $(\ker A)^{\perp} = \mathbb{R}^n$
(xviii) $\operatorname{row} A = \mathbb{R}^n$

(xix) $T(\mathbf{x}) = A\mathbf{x}$ is injective
(xx) $T(\mathbf{x}) = A\mathbf{x}$ is surjective
(xxi) $T(\mathbf{x}) = A\mathbf{x}$ is bijective, having $\operatorname{im} T = \mathbb{R}^n$ and an inverse T^{-1}

(xxii) A has n nonzero singular values
(xxiii) $\det A \neq 0$

2.2. DETERMINANTS

2.2.1. Permutations

Definition 2.2.1 (Permutation)

A permutation is a bijective map from a finite set to itself.

ⓘ Note

Permutations are commonly represented as value tables or graphs.

Definition 2.2.2 (Fixed Point)

A fixed point is invariant under a transformation such as a permutation.

Definition 2.2.3 (Symmetric Group)

The symmetric group of $X = \{1, \ldots, n\}$ is

$$S_n = \{\sigma : X \to X \mid \sigma \text{ is bijective}\}$$

and has $n!$ elements.

cyclic, e.g. (1 4 2) (3)

Definition 2.2.4 (Transposition)

A transposition of the elements $i, j \in \{1, \ldots, n\}$ is the permutation τ_{ij} with the cyclic notation (ij).

Theorem 2.2.1

A permutation $\sigma \in S_n$ can be written as a composition of transpositions.

Definition 2.2.5 (Sign Function)

The sign (or *signum*) of a number $x \in \mathbb{R}$ is

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ +1 & \text{if } x > 0, \end{cases}$$

while the sign of a permutation $\sigma \in S_n$ is

$$\operatorname{sgn} \sigma = \begin{cases} -1 & \text{if } \sigma \text{ is odd}, \\ +1 & \text{if } \sigma \text{ is even}, \end{cases}$$

or equivalently

$$\operatorname{sgn} \sigma = (-1)^k,$$

where k is the number of inversion or equivalently transpositions.

Example.

$$\frac{\mathrm{d}}{\mathrm{d}x}|x| = \operatorname{sgn}(x)$$

2.2.2. Computation

Definition 2.2.6 (Determinant)

The determinant of an $n \times n$ matrix A is the scalar

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Observe that elements from each row and column occur exactly once in every product.

ⓘ Note

The determinant measures, e.g.,

(i) whether a matrix is invertible,
(ii) the scaling factor of the linear transformation defined by A, and
(iii) how linearly independent the columns of A are.

Theorem 2.2.2 (Cofactor Expansion)

For $n > 1$,

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

where M_{1j} is the $(n-1) \times (n-1)$ minor obtained by deleting row 1 and column j from A.

Theorem 2.2.3 (Properties)

Let A and B be $n \times n$ matrices. Then

(i) $\det AB = (\det A)(\det B)$
(ii) $\det A^{\mathsf{T}} = \det A$
(iii) $\det A = a_{11}a_{22} \cdots a_{nn}$ if A is triangular

Theorem 2.2.4

The determinant is multilinear in its rows.

2.2.3. Linear Systems

$$\det A = \begin{cases} (-1)^r \left(\begin{array}{c}\text{product of} \\ \text{pivots in } U\end{array}\right) & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases}$$

Definition 2.2.7 (Cofactor Matrix Variant)

For any $n \times n$ matrix A and $\mathbf{b}$ with n elements we define

$$A_i(\mathbf{b}) = [\mathbf{a}_1 \ \cdots \ \mathbf{b} \ \cdots \ \mathbf{a}_n].$$

Lemma 2.2.5 (Cramer's Rule)

Let A be an invertible $n \times n$ matrix. For every $\mathbf{b} \in \mathbb{R}^n$ the unique solution has entries satisfying

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}.$$

PROOF Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If $A\mathbf{x} = \mathbf{b}$, the definition of matrix multiplication shows that

$$\begin{aligned} A \cdot I_i(\mathbf{x}) &= A\left[\mathbf{e}_1 \ \cdots \ \mathbf{x} \ \cdots \ \mathbf{e}_n\right] = \left[A\mathbf{e}_1 \ \cdots \ A\mathbf{x} \ \cdots \ A\mathbf{e}_n\right] \\ &= \left[\mathbf{a}_1 \ \cdots \ \mathbf{b} \ \cdots \ \mathbf{a}_n\right] = A_i(\mathbf{b}) \end{aligned}$$

By the multiplicative property of determinants,

$$(\det A)(\det I_i(\mathbf{x})) = \det A_i(\mathbf{b})$$

The second determinant on the left is simply x_i. (Make a cofactor expansion along the ith row.) Hence $(\det A) \cdot x_i = \det A_i(\mathbf{b})$. This proves (1) because A is invertible and $\det A \neq 0$. ■

2.2.4. Adjugates

Definition 2.2.8 (Adjugate)

The *adjugate* (or *classical adjoint*) $\operatorname{adj} A$ of a square matrix A is the transpose of its cofactor matrix.

Theorem 2.2.6

Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

Corollary 2.2.7

$$A \operatorname{adj} A = I \det A$$

2.2.5. Area and Volume

Theorem 2.2.8

If A is 2×2 then the area of the parallelogram determined by its columns is $|\det A|$.

If A is 3×3 then the volume of the parallelepiped determined by its columns is $|\det A|$.

Corollary 2.2.9

The area between $\mathbf{a}_1$ and $\mathbf{a}_2$ is the same as $\mathbf{a}_1$ and $\mathbf{a}_2 + \lambda\mathbf{a}_1$.

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\text{area of } T(S)\} = |\det A| \cdot \{\text{area of } S\} \tag{5}$$

If T is determined by a 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\text{volume of } T(S)\} = |\det A| \cdot \{\text{volume of } S\} \tag{6}$$

The theorem above holds for any region with finite area in $\mathbb{R}^2$ or finite volume in $\mathbb{R}^3$.

EXAMPLE 5 Let a and b be positive numbers. Find the area of the region E bounded by the ellipse whose equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

SOLUTION We claim that E is the image of the unit disk D under the linear transformation T determined by the matrix $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, because if $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathbf{x} = A\mathbf{u}$, then

$$u_1 = \frac{x_1}{a} \quad \text{and} \quad u_2 = \frac{x_2}{b}$$

It follows that $\mathbf{u}$ is in the unit disk, with $u_1^2 + u_2^2 \leq 1$, if and only if $\mathbf{x}$ is in E, with $(x_1/a)^2 + (x_2/b)^2 \leq 1$. By the generalization of Theorem 10,

$$\begin{aligned} \{\text{area of ellipse}\} &= \{\text{area of } T(D)\} \\ &= |\det A| \cdot \{\text{area of } D\} \\ &= ab\pi(1)^2 = \pi ab \end{aligned}$$

■

3. Linear Maps

3.1. DEFINITIONS

Definition 3.1.1 (Linear Map)

A map $T : V \to W$ is *linear* if

$$T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$$

for all $\mathbf{u}, \mathbf{v} \in V$ and $a, b \in \mathbb{R}$. In other words, the function satisfies additivity and homogeneity.

If $W = V$, then T is said to be an *operator*. Else if $W = F$, then T is said to be a *functional*.

Theorem 3.1.1

Any linear map T has an associated matrix $M(T)$.

Theorem 3.1.2

A linear map $T : V \to W$ is bijective if it is

(i) injective ($\ker T = \{\mathbf{0}\}$) and
(ii) surjective ($\operatorname{im} T = W$).

This is equivalent to T having a linear inverse $T^{-1} : W \to V$.

If V is finite-dimensional, then T being injective, surjective, or invertible is equivalent.

Definition 3.1.2 (Isomorphism)

A map $T : V \to W$ is an *isomorphism* if it is

(i) bijective and linear or, equivalently,
(ii) invertible.

If such a map exists, the spaces *isomorphic*, written $V \cong W$. This means that the difference between them is the choice of basis.

If $V = W$, then T is an *automorphism*.

Theorem 3.1.3

Two finite dimensional vectors spaces over F are isomorphic if and only if they have the same dimension.

Definition 3.1.3 (Affine)

A map f is affine if it can be written as

$$f(x) = Ax + b$$

or, equivalently, if the map g defined by

$$g(x) = f(x) - f(0)$$

is linear.

3.2. PROPERTIES

Theorem 3.2.1 (Left/Right Inverse)

If a linear map has a right inverse then it is surjective, if it has a left inverse then it is injective, and if it has both it is bijective.

3.3. IN VECTOR SPACES

Theorem 3.3.1

$\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$

Theorem 3.3.2

The set of all linear maps $\mathcal{L}(V, W)$ from V to W is a vector space.

Theorem 3.3.3 (Fundamental Theorem of Linear Maps)

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then, $\operatorname{im} T$ is finite-dimensional and

$$\dim \ker T + \dim \operatorname{im} T = \dim V.$$

4. Sequences

4.1. LINEAR

Theorem 4.1.1 (Linear Recurrences of Order Two)

Given

$$a_n = c_1 a_{n-1} + c_2 a_{n-2}$$

(i) Solve

$$r^2 - c_1 r - c_2 = 0$$

(ii) If distinct roots:

$$a_n = Ar_1^n + Br_2^n$$

If repeated roots:

$$a_n = (A + Bn)r^n$$

If complex roots:

$$a_n = R^n(A\cos(n\theta) + B\sin(n\theta)),$$

where $R = \sqrt{\alpha^2 + \beta^2}$ is the norm of $r = \alpha \pm i\beta$.

(iii) Determine constants A and B from a_1 and a_0.
(iv) For asymptotics find the root r of largest magnitude with some coefficient C and use that $a_n \sim Cr^n$ for large n to calculate

$$\lim_{n \to \infty} \frac{a_{n+1}}{a_n} = r.$$

That is, the ratio limit is the root of largest magnitude.

4.2. REPETITION OF REAL ANALYSIS

4.2.1. Definitions

Definition 4.2.1 (Sequence)

A *sequence* is a function whose domain is $\mathbb{N}$.

Definition 4.2.2 (Convergence)

A sequence *converges* to a if

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} : n \geq N \Longrightarrow |a_n - a| < \varepsilon$$

or equivalently if for any $V_\varepsilon(a)$ there exists a point in the sequence after which all terms are in $V_\varepsilon(a)$. In other words, if every ε-neighborhood of some point contains all but a finite number of the terms in (a_n).

We write this $\lim_{n \to \infty} a_n = \lim a_n = a$ or $a_n \to a$.

Example. Template of a typical convergence proof:

(i) Let $\varepsilon > 0$ be arbitrary.
(ii) Propose an $N \in \mathbb{N}$ (found before writing the proof).
(iii) Assume $n \geq N$.
(iv) Show that $|a_n - a| < \varepsilon$.

Theorem 4.2.1 (Uniqueness of Limits)

The limit of a sequence, if it exists, is unique.

4.2.2. Bounded Sequences

Definition 4.2.3 (Bounded)

A sequence is *bounded* if $\exists M > 0 : |a_n| < M \ \forall n \in \mathbb{N}$.

Theorem 4.2.2 (Convergent)

Every convergent sequence is bounded.

If a sequence is monotone and bounded it converges.

Subsequences of a convergent sequence converge to the same limit.

Theorem 4.2.3 (Bolzano–Weierstrass)

In a compact set $K \subset \mathbb{R}$, every bounded sequence contains a convergent subsequence whose limit point is in K.

4.2.3. Cauchy Sequences

Definition 4.2.4 (Cauchy Sequence)

A sequence (a_n) is a *Cauchy sequence* if

$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} : m, n \geq N \Longrightarrow |a_n - a_m| < \varepsilon.$$

Theorem 4.2.4 (Cauchy Criterion)

A sequence converges if and only if it is a Cauchy sequence.

5. Orthogonality

5.1. INNER PRODUCTS

Definition 5.1.1 (Inner Product)

The inner product for a real vector space V is function

$$\langle \mathbf{u}, \mathbf{v} \rangle : V \times V \to \mathbb{R}$$

with

(i) bilinearity
(ii) symmetry
(iii) positive-definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if $\mathbf{v} = \mathbf{0}$ else $\langle \mathbf{v}, \mathbf{v} \rangle > 0$

More generally for any vector space, the function must have

(i) linearity in the first argument
(ii) conjugate symmetry: $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
(iii) positive-definiteness

⚠ Warning

The *dot product* refers to the typical definition of the inner product.

Definition 5.1.2 (Inner Product Space)

An inner product space is a vector space and its inner product.

Examples.

$$\begin{aligned} &V = \mathbb{R}^n, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} \\ &V = \mathbb{R}^2, \quad \langle (x_1, x_2), (y_1, y_2) \rangle = 3x_1y_1 + 5x_2y_2 \\ &V = \mathbb{R}^2, \quad \langle (x_1, x_2), (y_1, y_2) \rangle = x_1y_1 + x_2y_2 + \frac{x_1y_2 + x_2y_1}{2} \\ &V = \mathbb{P}^n, \quad \langle \mathbf{p}, \mathbf{q} \rangle = p(t_0)q(t_0) + \cdots + p(t_n)q(t_n), \quad t_0 \neq \cdots \neq t_n \\ &V = C^0([-1, 1]), \quad \langle f, g \rangle = \frac{1}{b - a} \int_{-1}^{1} f(t)d(t)\,\mathrm{d}t \end{aligned}$$

Definition 5.1.3 (Norm)

In an inner product space, the norm is defined as

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

Definition 5.1.4 (Unit Vector)

A unit vector is one whose norm is 1.

Let V be a an inner product space over $\mathbb{R}$ or $\mathbb{C}$.

Lemma 5.1.1

Given a $\mathbf{v}$ in an product product space V and given a finite-dimensional subspace W, the Pythagorean Theorem in conjunction with the orthogonal decomposition of $\mathbf{v}$ with respect to W yields

$$\|\mathbf{v}\|^2 = \|\operatorname{proj}_W \mathbf{v}\|^2 + \|\mathbf{v} - \operatorname{proj}_W \mathbf{v}\|^2.$$

In particular, note that

$$\|\mathbf{v}\| \geq \|\operatorname{proj}_W \mathbf{v}\|.$$

Theorem 5.1.2 (The Cauchy–Schwarz Inequality)

For all $\mathbf{u}, \mathbf{v} \in V$,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Theorem 5.1.3 (The Triangle Inequality)

For all $\mathbf{u}, \mathbf{v} \in V$,

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

5.2. VECTORS

Definition 5.2.1 (Orthogonality)

The vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner product space are orthogonal, written $x \perp y$, if and only if $\langle x, y \rangle = 0$.

ⓘ Note

The zero-vector is orthogonal to all vectors.

Definition 5.2.2

Planes are said to be orthogonal if their normal vectors are orthogonal.

5.3. PROJECTION

5.3.1. Intersections

💡 Finding Intersections

- line–plane
  Substitute the line's parametric form into the plane's general equation.
- line–line
  Solve the system formed by the two parametric equations.
- plane–plane
  Construct a vector between two arbitrary points (one on each plane) given in parametric form. Set its dot product with each plane's normal vector to zero. Solve the resulting system of equations.

5.3.2. Orthogonal Basis

Definition 5.3.1 (Orthogonal Projection)

The projection of $\mathbf{y}$ onto W is the $\hat{\mathbf{y}}$ in W such that

$$\mathbf{y} - \hat{\mathbf{y}} \perp W.$$

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \qquad (j = 1, \ldots, p)$$

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\operatorname{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_p]$, then

$$\operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U. Then

$$\begin{aligned} U^TU\mathbf{x} &= I_p\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^p \qquad \text{Theorem 6} \\ UU^T\mathbf{y} &= \operatorname{proj}_W \mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \qquad \text{Theorem 10} \end{aligned}$$

If U is an $n \times n$ (square) matrix with orthonormal columns, then U is an *orthogonal* matrix, the column space W is all of $\mathbb{R}^n$, and $UU^T\mathbf{y} = I\mathbf{y} = \mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1} \end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \qquad \text{for } 1 \leq k \leq p$$

The QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

💡 Tip

To find R during QR-factorization, find Q and set

$$R = Q^{\mathsf{T}}A.$$

5.3.3. Approximation

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

⚠ Warning

The least-squares error is sometimes defined as the squared norm.

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
b. The columns of A are linearly independent.
c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Theorem 5.3.1

If $\mathbf{b} \perp \operatorname{col} A$ then $\hat{\mathbf{b}} := \operatorname{proj}_{\operatorname{col} A} \mathbf{b} = \mathbf{0}$.

5.4. SETS

Definition 5.4.1 (Orthogonal Complement)

The orthogonal complement to a subset S of a vector space V is

$$S^{\perp} = \{\mathbf{x} \in V : \mathbf{x} \perp \mathbf{s} \ \forall \mathbf{s} \in S\}.$$

Definition 5.4.2 (Orthogonal Set)

A set $S \subseteq V$ is orthogonal if $\mathbf{x} \perp \mathbf{y}$ for every $\mathbf{x} \neq \mathbf{y} \in S$. The set is also orthonormal (ON) if $\mathbf{x}$ for every $\mathbf{x} \in S$.

Theorem 5.4.1

Orthogonal sets of nonzero vectors are linearly independent.

`proof`

5.5. MATRICES

Definition 5.5.1 (Orthogonal Matrix)

An *orthogonal* (also called *orthonormal*) matrix is an invertible square matrix Q with $Q^{\mathsf{T}} = Q^{-1}$.

Such a matrix has orthonormal rows and columns.

Theorem 5.5.1

An $m \times n$ matrix U has orthonormal columns if and only if $U^{\mathsf{T}}U = I_n$.

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
b. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
c. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

Theorem 5.5.2

Let $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ be an ON-basis to an inner product space V of which $\mathbf{x} \in V$. Then,

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \langle \mathbf{x}, \mathbf{b}_1 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{b}_n \rangle \end{bmatrix}.$$

`proof`

Theorem 5.5.3 (Orthogonal Complements)

For any matrix A,

$$(\operatorname{row} A)^{\perp} = \ker A \quad \text{and} \quad (\operatorname{col} A)^{\perp} = \operatorname{coker} A.$$

6. Eigenvalues and Eigenvectors

6.1. OF MATRICES

Definition 6.1.1 (Eigenstuff)

An eigenvalue λ of matrix $A \in \mathbb{C}^{n \times n}$ satisfies

$$A\mathbf{x} = \lambda\mathbf{x} \text{ or, equivalently, } (A - \lambda I)\mathbf{x} = \mathbf{0}$$

for some eigenvector $\mathbf{x} \neq \mathbf{0}$. The eigenvalues are found from the characteristic equation

$$\det(A - \lambda I) = 0,$$

which for each eigenvalue gives the eigenvectors as the solutions to

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

and the eigenspace as

$$E_\lambda = \ker(A - \lambda I),$$

which is a subspace of $\mathbb{C}^n$.

The multiplicity of a solution to the characteristic equation is the *algebraic multiplicity* of that eigenvalue.

ⓘ Note

Although the eigenvector must be nonzero, the eigenvalue may be zero, which is it if and only if the matrix is not invertible.

Proof. Since 0 is an eigenvalue, $0 = \det(A - 0I)$, but $A - 0I = A$. □

Theorem 6.1.1

The eigenvalues of a triangular matrix are the entries on its diagonal.

Let λ be an eigenvalue of a triangular matrix A. Then, if and only if $a_{ii} = \lambda$ for some diagonal element a_{ii}, that gives a free

Theorem 6.1.2

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix, then they are linearly independent.

Definition 6.1.2 (Trace)

The *trace* of an $n \times n$ matrix $A = (a_{ij})$ is the sum of its diagonal entries:

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

Basic properties:

(i) (Linearity) $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$ and $\operatorname{tr}(cA) = c \operatorname{tr} A$.
(ii) (Cyclic) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ (when both products are defined).
(iii) (Similarity) $\operatorname{tr}(P^{-1}AP) = \operatorname{tr} A$ for invertible P.

6.2. REAL

6.2.1. Diagonalization

Definition 6.2.1 (Similarity)

If A and B are $n \times n$ matrices such t hat there exists P with

$$P^{-1}AP = B \text{ or, equivalently, } A = PBP^{-1}.$$

Theorem 6.2.1

If two matrices are similar, then they share eigenvalues.

However, the reverse is not implied.

Theorem 6.2.2 (Diagonalization)

For an $n \times n$ matrix A it is equivalent that

(i) A is diagonalizable.
(ii) A has n linearly independent eigenvectors.
(iii) There exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$. In fact, P is the eigenvectors and D is the eigenvalues.

Corollary 6.2.3

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

💡 Tip

To avoid computing P^{-1}, simply verify that $AP = PD$. This is equivalent to $A = PDP^{-1}$ if P is invertible (which should be verified).

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p$.

a. For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.
b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (i) the characteristic polynomial factors completely into linear factors and (ii) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.
c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \ldots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

6.2.2. Maps

Definition 6.2.2 (Eigenvalues and Eigenvectors of a Map)

Let V be a vector space. An *eigenvector* of a linear map $T : V \to V$ is a nonzero vector $\mathbf{x} \in V : T(\mathbf{x}) = \lambda\mathbf{x}$ for some scalar λ.

A scalar λ is called an *eigenvalue* of T if there is a nontrivial solution $\mathbf{x}$ to $T(\mathbf{x}) = \lambda\mathbf{x}$.

Such an x is called an *eigenvector* corresponding to λ.

Diagonal Matrix Representation

Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

PROOF Denote the columns of P by $\mathbf{b}_1, \ldots, \mathbf{b}_n$, so that $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $P = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$. In this case, P is the change-of-coordinates matrix $P_{\mathcal{B}}$ discussed in Section 4.4, where

$$P[\mathbf{x}]_{\mathcal{B}} = \mathbf{x} \quad \text{and} \quad [\mathbf{x}]_{\mathcal{B}} = P^{-1}\mathbf{x}$$

If $T(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^n$, then

$$\begin{aligned} [T]_{\mathcal{B}} &= \left[[T(\mathbf{b}_1)]_{\mathcal{B}} \ \cdots \ [T(\mathbf{b}_n)]_{\mathcal{B}}\right] && \text{Definition of } [T]_{\mathcal{B}} \\ &= \left[[A\mathbf{b}_1]_{\mathcal{B}} \ \cdots \ [A\mathbf{b}_n]_{\mathcal{B}}\right] && \text{Since } T(\mathbf{x}) = A\mathbf{x} \\ &= \left[P^{-1}A\mathbf{b}_1 \ \cdots \ P^{-1}A\mathbf{b}_n\right] && \text{Change of coordinates} \\ &= P^{-1}A[\mathbf{b}_1 \ \cdots \ \mathbf{b}_n] && \text{Matrix multiplication} \\ &= P^{-1}AP \end{aligned} \tag{6}$$

Since $A = PDP^{-1}$, we have $[T]_{\mathcal{B}} = P^{-1}AP = D$. ■

6.2.3. The Real Spectral Theorem

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

PROOF Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, compute

$$\begin{aligned} \lambda_1\mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1\mathbf{v}_1)^T\mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2 && \text{Since } \mathbf{v}_1 \text{ is an eigenvector} \\ &= (\mathbf{v}_1^TA^T)\mathbf{v}_2 = \mathbf{v}_1^T(A\mathbf{v}_2) && \text{Since } A^T = A \\ &= \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) && \text{Since } \mathbf{v}_2 \text{ is an eigenvector} \\ &= \lambda_2\mathbf{v}_1^T\mathbf{v}_2 = \lambda_2\mathbf{v}_1 \cdot \mathbf{v}_2 \end{aligned}$$

Hence $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But $\lambda_1 - \lambda_2 \neq 0$, so $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. ■

The Spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.
b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.
c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.
d. A is orthogonally diagonalizable.

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T \tag{2}$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. Each term in (2) is an $n \times n$ matrix of rank 1. For example, every column of $\lambda_1\mathbf{u}_1\mathbf{u}_1^T$ is a multiple of $\mathbf{u}_1$. Furthermore, each matrix $\mathbf{u}_j\mathbf{u}_j^T$ is a **projection matrix** in the sense that for each $\mathbf{x}$ in $\mathbb{R}^n$, the vector $(\mathbf{u}_j\mathbf{u}_j^T)\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{u}_j$. (See Exercise 41.)

6.3. QUADRATIC FORMS

6.3.1. Definition

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

The simplest example of a nonzero quadratic form is $Q(\mathbf{x}) = \mathbf{x}^TI\mathbf{x} = \|\mathbf{x}\|^2$. Examples 1 and 2 show the connection between any symmetric matrix A and the quadratic form $\mathbf{x}^TA\mathbf{x}$.

Important

Any symmetric matrix has a valid quadratic form.

In some cases, quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix. Fortunately, the cross-product term can be eliminated by making a suitable change of variable.

Change of Variable in a Quadratic Form

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\mathbf{x} = P\mathbf{y}, \quad \text{or equivalently,} \quad \mathbf{y} = P^{-1}\mathbf{x} \tag{1}$$

where P is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. Here $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the basis of $\mathbb{R}^n$ determined by the columns of P. (See Section 4.4.)

If the change of variable (1) is made in a quadratic form $\mathbf{x}^TA\mathbf{x}$, then

$$\mathbf{x}^TA\mathbf{x} = (P\mathbf{y})^TA(P\mathbf{y}) = \mathbf{y}^TP^TAP\mathbf{y} = \mathbf{y}^T(P^TAP)\mathbf{y} \tag{2}$$

and the new matrix of the quadratic form is P^TAP. Since A is symmetric, Theorem 2 guarantees that there is an orthogonal matrix P such that P^TAP is a diagonal matrix D, and the quadratic form in (2) becomes $\mathbf{y}^TD\mathbf{y}$. This is the strategy of the next example.

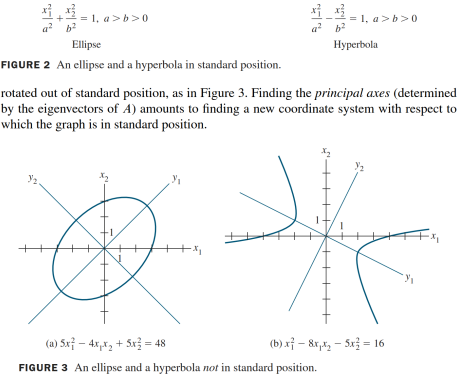
6.3.2. Principal Axes

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^TA\mathbf{x}$ into a quadratic form $\mathbf{y}^TD\mathbf{y}$ with no cross-product term.

The columns of P in the theorem are called the **principal axes** of the quadratic form $\mathbf{x}^TA\mathbf{x}$. The vector $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the orthonormal basis of $\mathbb{R}^n$ given by these principal axes.

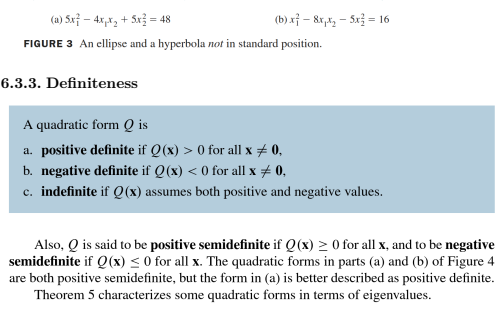
point, or contains no points at all. If A is a diagonal matrix, the graph is in *standard position*, such as in Figure 2. If A is not a diagonal matrix, the graph of equation (3) is



x_2
b
a
x_1
$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, \; a > b > 0$
Ellipse
x_2
b
a
x_1
$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1, \; a > b > 0$
Hyperbola

FIGURE 2 An ellipse and a hyperbola in standard position.

rotated out of standard position, as in Figure 3. Finding the *principal axes* (determined by the eigenvectors of A) amounts to finding a new coordinate system with respect to which the graph is in standard position.



x_2
y_2
y_1
x_1
(a) $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$
x_2
y_2
y_1
x_1
(b) $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

FIGURE 3 An ellipse and a hyperbola *not* in standard position.

6.3.3. Definiteness

A quadratic form Q is

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,
b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,
c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.

Also, Q is said to be **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, and to be **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$. The quadratic forms in parts (a) and (b) of Figure 4 are both positive semidefinite, but the form in (a) is better described as positive definite.

Theorem 5 characterizes some quadratic forms in terms of eigenvalues.

Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^TA\mathbf{x}$ is

a. positive definite if and only if the eigenvalues of A are all positive,
b. negative definite if and only if the eigenvalues of A are all negative, or
c. indefinite if and only if A has both positive and negative eigenvalues.

PROOF By the Principal Axes Theorem, there exists an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = \mathbf{y}^TD\mathbf{y} = \lambda_1y_1^2 + \lambda_2y_2^2 + \cdots + \lambda_ny_n^2 \tag{4}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Since P is invertible, there is a one-to-one correspondence between all nonzero $\mathbf{x}$ and all nonzero $\mathbf{y}$. Thus the values of $Q(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$ coincide with the values of the expression on the right side of (4), which is obviously controlled by the signs of the eigenvalues $\lambda_1, \ldots, \lambda_n$, in the three ways described in the theorem. ■

The classification of a quadratic form is often carried over to the matrix of the form. Thus a **positive definite matrix** A is a *symmetric* matrix for which the quadratic form $\mathbf{x}^TA\mathbf{x}$ is positive definite. Other terms, such as **positive semidefinite matrix**, are defined analogously.

6.4. COMPLEX

6.4.1. Complex Calculations

Definition 6.4.1 (Complex Vector)

$$\mathbf{x} = \operatorname{Re}\mathbf{x} + i\operatorname{Im}\mathbf{x}$$

$$\bar{\mathbf{x}} = \operatorname{Re}\mathbf{x} - i\operatorname{Im}\mathbf{x}$$

If B is an $m \times n$ matrix with possibly complex entries, then $\bar{B}$ denotes the matrix whose entries are the complex conjugates of the entries in B. Let r be a complex number and $\mathbf{x}$ any vector. Properties of conjugates for complex numbers carry over to complex matrix algebra:

$$\overline{r\mathbf{x}} = \bar{r}\bar{\mathbf{x}}, \quad \overline{B\mathbf{x}} = \bar{B}\bar{\mathbf{x}}, \quad \overline{BC} = \bar{B}\bar{C}, \quad \text{and} \quad \overline{rB} = \bar{r}\bar{B}$$

Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and an associated eigenvector $\mathbf{v}$ in $\mathbb{C}^2$. Then

$$A = PCP^{-1}, \quad \text{where} \quad P = [\operatorname{Re}\mathbf{v} \ \ \operatorname{Im}\mathbf{v}] \quad \text{and} \quad C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

EXAMPLE 3 One way to see how multiplication by the matrix A in Example 2 affects points is to plot an arbitrary initial point—say, $\mathbf{x}_0 = (2, 0)$—and then to plot successive images of this point under repeated multiplications by A. That is, plot

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} \\ \mathbf{x}_2 &= A\mathbf{x}_1 = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}\begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} = \begin{bmatrix} -.4 \\ 2.4 \end{bmatrix} \\ \mathbf{x}_3 &= A\mathbf{x}_2, \ldots \end{aligned}$$

Figure 1 shows $\mathbf{x}_0, \ldots, \mathbf{x}_8$ as larger dots. The smaller dots are the locations of $\mathbf{x}_9, \ldots, \mathbf{x}_{100}$. The sequence lies along an elliptical orbit. ■

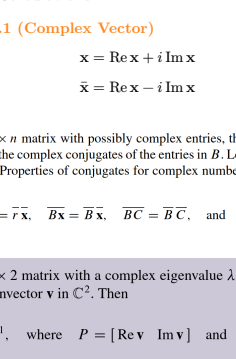


FIGURE 1 Iterates of a point $\mathbf{x}_0$ under the action of a matrix with a complex eigenvalue.

Of course, Figure 1 does not explain *why* the rotation occurs. The secret to the rotation is hidden in the real and imaginary parts of a complex eigenvector.

6.4.2. Hermitian Spaces

Definition 6.4.2 (Hermitian Form)

A Hermitian form $\langle -, - \rangle : V \times V \to \mathbb{C}$ is a bilinear form on a complex vector space with the following properties:

(i) $\mathbb{R}$ bilinearity, i.e.,
- $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$
- $\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w} \rangle = \bar{a}\langle \mathbf{u}, \mathbf{v} \rangle + \bar{b}\langle \mathbf{u}, \mathbf{w} \rangle$

(ii) conjugate symmetry: $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
(iii) positive-definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if $\mathbf{v} = \mathbf{0}$ else $\langle \mathbf{v}, \mathbf{v} \rangle > 0$
(iv) $\langle i\mathbf{v}, \mathbf{w} \rangle = i\langle \mathbf{v}, \mathbf{w} \rangle = -\langle \mathbf{v}, i\mathbf{w} \rangle$

Example. The Hermitian standard form on $\mathbb{C}^n$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \bar{\mathbf{y}} = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n.$$

Definition 6.4.3 (Hermitian Space)

A Hermitian space is a $\mathbb{C}$ vector space with a Hermitian form.

ⓘ Note

The complex transpose (also called Hermitian transpose) is notated A^*, A^{H}, A' or $A^\dagger$.

Theorem 6.4.1

For the Hermitian standard form on $\mathbb{C}^n$ and $A \in \mathbb{C}^{n \times n}$ it holds that

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^*\mathbf{y} \rangle.$$

Definition 6.4.4

A matrix A is Hermitian if $A = A^*$.

Theorem 6.4.2

All Hermitian bilinear forms can be written

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^*A\mathbf{y}$$

for some Hermitian A.

Corollary 6.4.3

All inner products can be written

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathsf{T}}A\mathbf{y}$$

for some symmetric A.

Definition 6.4.5 (Hermitian Map)

Let $(V, \langle -, - \rangle)$ be a Hermitian space. A $\mathbb{C}$-linear map $f : V \to V$ is Hermitian if

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

This is similar to an $\mathbb{R}$-linear map f in an inner product space being symmetric.

Theorem 6.4.4

Let $\mathcal{B}$ be a unitary basis to a Hermitian room V. Then f is Hermitian if and only if $A = [f]_{\mathcal{B}}$ is Hermitian.

6.4.3. Unitary Bases

Definition 6.4.6 (Kronecker Delta)

The Kronecker delta is

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

and is used to express orthonormality and identity matrices.

Definition 6.4.7 (Unitary Basis)

A unitary basis to a Hermitian space is a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}.$$

Definition 6.4.8 (Unitary Matrix)

A matrix $U \in \mathbb{C}^{n \times n}$ is unitary if its columns form a unitary basis to $\mathbb{C}^n$ or, equivalently, if $UU^* = I$ or $U^{-1} = U^*$.

ⓘ Note

A symmetric matrix is a real Hermitian matrix.

An orthonormal basis is a real unitary basis.

An orthogonal matrix is a real unitary matrix.

6.4.4. The Complex Spectral Theorem

Definition 6.4.9 (Unitary Diagonalization)

A matrix $A \in \mathbb{C}^{n \times n}$ is unitary diagonalizable if there exists a *real* diagonal matrix D and a unitary matrix U such that

$$A = UDU^{-1} = UDU^*.$$

A map of complex vector spaces is unitary diagonalizable if there exists a unitary basis $\mathcal{B}$ such that $[f]_{\mathcal{B}}$ is a real diagonal matrix.

Lemma 6.4.5

All eigenvalues to Hermitian maps of finite-dimensional Hermitian spaces are real.

Proof. Let $A : V \to V$ be Hermitian and remark that

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle \quad \forall \mathbf{x}, \mathbf{y} \in V.$$

Now if $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, then

$$\lambda\langle \mathbf{v}, \mathbf{v} \rangle = \langle A\mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, A\mathbf{v} \rangle = \bar{\lambda}\langle \mathbf{v}, \mathbf{v} \rangle.$$

Since $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ it follows that $\lambda = \bar{\lambda}$, hence $\lambda \in \mathbb{R}$. □

Theorem 6.4.6 (The Complex Spectral Theorem)

A Hermitian map of a finite-dimensional $\mathbb{C}$ vector space is unitary diagonalizable.

Corollary 6.4.7

A Hermitian matrix $A \in \mathbb{C}^{n \times n}$ is unitary diagonalizable. That is,

$$A = A^* \Leftrightarrow A = UDU^*.$$

7. Statistics

7.1. LINEAR REGRESSION

$X\boldsymbol{\beta} = \mathbf{y}$, where X is the design matrix, $\boldsymbol{\beta}$ is the parameter vector of regression coefficients, and $\mathbf{y}$ is the observation vector.

The simplest relation between two variables is the linear equation $y = \beta_0 + \beta_1x$. The difference between predicted and observed y-value is called a residual.

If written on mean-deviation form $x^* = x - \bar{x}$ then the two columns of the design matrix will be orthogonal.

A linear model is on the form

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where X is the design matrix, $\boldsymbol{\beta}$ is the parameter vector of regression coefficients, and $\mathbf{y}$ is the observation vector. Its solution solution $\hat{\boldsymbol{\beta}}$ satisfies

$$X^{\mathsf{T}}X\boldsymbol{\beta} = X^{\mathsf{T}}\mathbf{y}.$$

$$\underset{\text{Observation vector}}{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad \underset{\text{Design matrix}}{X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}}, \quad \underset{\text{Parameter vector}}{\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}, \quad \underset{\text{Residual vector}}{\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

7.2. MARKOV CHAINS

Definition 7.2.1 (Markov Chain)

A Markov chain is a sequence of random variables $X_1, X_2, \ldots$. The process is said to have the Markov property if

$$P(X_{n+1} = x \mid X_n = x_n, \ldots, X_1 = x_1) = P(X_{n+1} = x \mid X_n = x_n)$$

for all n and all states $x_1, \ldots, x_n, x$.

A Markov chain is characterized by its transition matrix P, where

$$P_{ij} = P(X_{n+1} = j \mid X_n = i).$$

Theorem 7.2.1

Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ be a probability vector. Then, $\boldsymbol{\pi}$ is a stationary distribution of the Markov chain with transition matrix P if and only if

$$\boldsymbol{\pi}^{\mathsf{T}}P = \boldsymbol{\pi}^{\mathsf{T}}.$$

Definition 7.2.2 (Distributions)

A *probability vector* (or *distribution*) on S is a row vector

$$\boldsymbol{\pi} = (\pi_1, \cdots, \pi_n)$$

with $\pi_i \geq 0$ and $\pi_1 + \cdots + \pi_n = 1$.

If $\boldsymbol{\pi}^0$ is the initial distribution of X_0, then the distribution after k steps is

$$\boldsymbol{\pi}^k = \boldsymbol{\pi}^0P^k.$$