Introduction to Linear Algebra

Elijah Kin

May 6, 2025

Contents

1 Matrix Algebra

1.1 Matrix Multiplication

Definition 1.1.1. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then AB denotes the $m \times p$ matrix with entries given by

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Note that this definition implies that in order for the product AB to be defined, the number of columns of A must match the number of rows of B.

Exercise 1.1.2. Given the matrices

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -5 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & -5 & 1 \\ 1 & -4 & -3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}, \quad E = \begin{bmatrix} -5 \\ 3 \end{bmatrix},$$

compute each of the following expressions:

$$-2A, \quad B - 2A, \quad AC, \quad CD, \quad A + 2B, \quad 3C - E, \quad CB, \quad EB$$

If an expression is undefined, explain why.

Exercise 1.1.3. If a matrix A is 5×3 and the product AB is 5×7, what is the size of B?

Proposition 1.1.4. Let A, B, and C be matrices such that the sums and products below are defined. It is not difficult to prove the following properties of matrix multiplication.

(a) Associativity: $A(BC) = (AB)C$

(b) Left Distributivity: $A(B + C) = AB + AC$

(c) Right Distributivity: $(B + C)A = BA + CA$

Matrix multiplication is not commutative in general; indeed, we will see examples of matrices A and B such that $AB \neq BA$.

Exercise 1.1.5. Given the matrices

$$A = \begin{bmatrix} 2 & 5 \\ -3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -5 \\ 3 & k \end{bmatrix}$$

what value(s) of k, if any, will make $AB = BA$?

Exercise 1.1.6. Given the matrices

$$A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$$

verify that $AB = AC$ and yet $B \neq C$. This shows that matrix multiplication does not satisfy "cancellation" either.

1.2 Matrix Transpose

Definition 1.2.1. If A is an $m \times n$ matrix, then A^T denotes the $n \times m$ matrix with entries

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

which we call the **transpose** of A.

Proposition 1.2.2. If A and B are matrices such that the following are defined, then

(a) $(A^T)^T = A$

(b) $(A+B)^T = A^T + B^T$

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

1.3 Matrix Inverse

Definition 1.3.1. We write I_n to denote the $n \times n$ **identity matrix**, whose entries are given by

$$(I_n)_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

To give a few concrete examples,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

When n is clear from context, we may also simply denote it as I.

Definition 1.3.2. If A is an $n \times n$ matrix, it is said to be **invertible** if there exists an $n \times n$ matrix B such that

$$BA = I_n = AB$$

in which case B is said to be the **inverse** of A. The usage of the word *the* here is justified, since the matrix inverse is unique; suppose A had another inverse C. Then

$$B = BI_n = BAC = I_nC = C,$$

showing the inverse is unique. Hence, we are justified in writing A^{-1} to denote the inverse of A. A matrix that is not invertible is sometimes called a **singular** matrix.

Proposition 1.3.3. If A and B are invertible matrices such that AB is defined, then

(a) $(A^{-1})^{-1} = A$

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

(c) $(A^T)^{-1} = (A^{-1})^T$

Proposition 1.3.4. In practice, we can invert a matrix via an algorithm involving row operations. In particular, we can row reduce the augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$. If A is row equivalent to I, then $\begin{bmatrix} A & I \end{bmatrix}$ is row equivalent to $\begin{bmatrix} I & A^{-1} \end{bmatrix}$. Otherwise, A does not have an inverse.

Exercise 1.3.5. Find the inverse of the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$$

if it exists.

Proposition 1.3.6. 2×2 matrices are small enough such that it is convenient to write an explicit formula for their inverse. In particular, suppose

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then as long as $ad - bc \neq 0$, A is invertible, and its inverse is

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The quantity $ad - bc$ is called the **determinant** of A. We will see much more of it in the future.
Exercise 1.3.7. Suppose A, B, and X are $n \times n$ matrices with A, X, and $A - AX$ invertible, and suppose

$$(A - AX)^{-1} = X^{-1}B.$$

(a) Explain why B is invertible.

(b) Solve the above equation for X. If you need to invert a matrix, explain why that matrix is invertible

Exercise 1.3.8. If A, B, and C are $n \times n$ invertible matrices, does the equation

$$C^{-1}(A + X)B^{-1} = I_n$$

have a solution, X? If so, find it.

2 Determinants

2.1 Cofactor Expansion

Definition 2.1.1. Let $n \geq 2$ and let A be an $n \times n$ matrix. The **determinant** of A is the quantity

$$\det(A) = \sum_{j=1}^{N} (-1)^{1+j} a_{1j} A_{1j}$$

where A_{1j} is the $(n-1) \times (n-1)$ matrix obtained by deleting the 1st row and the jth column from A. Note that the values that the a_{1j} term takes over this summation are the entries in the first row of A; hence, this formula is known as the **cofactor expansion** across the first row of A. It can be proven that analogous cofactor expansions across any row or column of A will yield the same quantity.
Exercise 2.1.2. Compute the determinant of

$$A = \begin{bmatrix} 4 & 0 & 0 & 5 \\ 1 & 7 & 2 & -5 \\ 3 & 0 & 0 & 0 \\ 8 & 3 & 1 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 0 & 2 & 4 & 0 \\ 9 & 0 & -4 & 1 & 0 \\ 8 & -5 & 6 & 7 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 3 & 2 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 7 & -2 & 0 & 0 \\ 2 & 6 & 3 & 0 \\ 3 & -8 & 4 & -3 \end{bmatrix}$$

via cofactor expansion. With careful choice of the row or column at each step, this requires minimal computation.

Solution: The natural choice is to cofactor expand across the third row, seeing as it is primarily zeros. Continuing this strategy of cofactor expanding across the row or column with the most zeros at each step, we find

$$\det(A) = 3\det\begin{bmatrix}0 & 0 & 5\\ 7 & 2 & -5\\ 3 & 1 & 7\end{bmatrix} = 3\cdot 5\det\begin{bmatrix}7 & 2\\ 3 & 1\end{bmatrix} = 3\cdot 5\cdot 1 = 15.$$

Proposition 2.1.3. An $n\times n$ matrix A is invertible if and only if $\det(A)\neq 0$.

2.2 A Cheaper Algorithm

Remark 2.2.1. We have seen that we can obtain useful information about a matrix from its determinant. For example, to answer whether A is invertible, we can simply check that $\det(A)\neq 0$.

However, computing determinants by cofactor expansion is wildly inefficient; cofactor expansion of an arbitrary $n\times n$ matrix requires on the order of $n!$ multiplications. When $n=3$ or 4, this is not so bad, but for $n=25$, a computer performing one trillion multiplications per second would still take about 500,000 years. This motivates our search for alternative means of computation.

Proposition 2.2.2. The determinant plays nicely with elementary row operations. In particular,

(a) If a multiple of one row of A is added to another to produce a matrix B, then $\det B = \det A$.

(b) If two rows of A are interchanged to produce a matrix B, then $\det B = -\det A$.

(c) If one row of A is multiplied by k to produce B, then $\det B = k\det A$.

This proposition will serve as the basis for our more efficient determinant computation. In particular, given a matrix A, if we can apply row operations until we reach some matrix B for which we can compute $\det B$ cheaply, this proposition will allow us to recover $\det A$. This begs the question, for which matrices B can we compute $\det B$ cheaply?

Definition 2.2.3. Let A be an $n\times n$ matrix. If all entries above the main diagonal of A are 0, then we say that A is **lower triangular**. Likewise, if all entries below the main diagonal are 0, then we say it is **upper triangular**.

The matrices L and U below are examples of $n\times n$ lower triangular and upper triangular matrices respectively. Note that our definition allows non-zero entries to lie on the diagonal.

$$L = \begin{bmatrix}\ell_{11} & 0 & 0 & \dots & 0\\ \ell_{21} & \ell_{22} & 0 & \dots & 0\\ \ell_{31} & \ell_{32} & \ell_{33} & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \ell_{n1} & \ell_{n2} & \ell_{n3} & \dots & \ell_{nn}\end{bmatrix}, \quad U = \begin{bmatrix}u_{11} & u_{12} & u_{13} & \dots & u_{1n}\\ 0 & u_{22} & u_{23} & \dots & u_{22}\\ 0 & 0 & u_{33} & \dots & u_{3n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & u_{nn}\end{bmatrix}$$

A **triangular** matrix is simply one which is either lower triangular or upper triangular.

Proposition 2.2.4. If A is an $n\times n$ triangular matrix, then

$$\det A = \prod_{k=1}^{n} a_{kk}$$

That is, for triangular matrices, the determinant is the product of its diagonal entries. No cofactor expansion required!
Example 2.2.5. Find the determinants of the below matrices

$$A = \begin{bmatrix} 1 & 5 & -4 \\ -1 & -4 & 5 \\ -2 & -8 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 & 0 & 2 \\ -2 & -5 & 7 & 4 \\ 3 & 5 & 2 & 1 \\ 1 & -1 & 2 & -3 \end{bmatrix}$$

by row reduction to echelon form.

2.3 Properties of the Determinant

Remark 2.3.1. Another way to cheapen the cost of computing determinants is to put the determinant of the expression of interest into terms already known determinants. This section will explore properties of the determinant that make this possible.
Exercise 2.3.2. Suppose we have the matrix

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

Is it true that $\det 5A = 5 \det A$? Further, now let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and let k be a scalar. Find a formula that relates $\det kA$ to k and $\det A$.
Proposition 2.3.3. Let A and B be $n \times n$ matrices. Then

(a) $\det A^T = \det A$

(b) $\det AB = \det(A)\det(B)$

Note that property (b) implies several other properties; for example, $\det(A^k) = \det(A)^k$ and $\det(kA) = k^n \det(A)$. Try to convince yourself that these properties follow.
Exercise 2.3.4. Let A and B be 3×3 matrices, with $\det A = -2$ and $\det B = 3$. Use properties of determinants to compute:

$$\det AB, \quad \det 5A, \quad \det B^T, \quad \det A^{-1}, \quad \det A^3$$

Exercise 2.3.5. Let A and B be 4×4 matrices, with $\det A = 4$ and $\det B = -3$. Compute:

$$\det AB, \quad \det B^5, \quad \det 2A, \quad \det A^T BA, \quad \det B^{-1}AB$$

2.4 Cramer's Rule

Proposition 2.4.1 (Cramer's Rule)**.** Let A be an $n \times n$ invertible matrix. Then for any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \quad i = 1, 2, \ldots, n$$

where $A_i(\mathbf{b})$ denotes the $n \times n$ matrix obtained from A by replacing column i by the vector $\mathbf{b}$.
Exercise 2.4.2. Use Cramer's rule to compute the solutions of the system

$$\begin{aligned} x_1 + 3x_2 + x_3 &= 8 \\ -x_1 + 2x_3 &= 4 \\ 3x_1 + x_2 &= 4 \end{aligned}$$

3 Abstract Vector Spaces

Remark 3.0.1. Until now, we have primarily considered vectors only as elements of $\mathbb{R}^n$, that is, as lists of real numbers. We will now broaden the meaning of "vector" and reconsider many previously-studied concepts using this new definition.

3.1 Vector Spaces

Definition 3.1.1. A **vector space** (over a field of scalars[1] $\mathbb{F}$) is a nonempty set V together with two operations $+$ and $\cdot$ such that all the following are true for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $c, d \in \mathbb{F}$.

(a) V is closed under addition: $\mathbf{u} + \mathbf{v} \in V$.

(b) Addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

(c) Addition is associative: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

(d) There exists a **zero vector**, notated $\mathbf{0}$, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.

(e) V is closed under additive inverses: there exists a $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

(f) V is closed under scalar multiplication: $c\mathbf{u} \in V$.

(g) Scalar multiplication distributes over addition: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.

(h) Scalar multiplication distributes over scalar addition: $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$.

(i) Scalar multiplication is compatible with field multiplication: $c(d\mathbf{u}) = (cd)\mathbf{u}$.

(j) The scalar 1 is the identity element of scalar multiplication: $1\mathbf{u} = \mathbf{u}$.

If V is a vector space, we refer to the elements of V as **vectors**.
Example 3.1.2. $\mathbb{R}^n$ with $n \geq 1$ is a vectors space, and what we have primarily studied until now.
Example 3.1.3. Let $n \geq 0$ and define

$$\begin{aligned}\mathbb{P}_n &= \{p(t) : p(t) \text{ is a polynomial of degree at most } n\} \\ &= \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n : a_0, a_1, a_2, \ldots, a_n \in \mathbb{R}\}.\end{aligned}$$

We claim that $\mathbb{P}_n$ is a vector space with addition and scalar multiplication defined in the "natural" way. What would the zero vector in this vector space be?

3.2 Subspaces

Definition 3.2.1. Let V be a vector space. A subset $H \subseteq V$ is a **subspace** of V if H is also a vector space (with the same addition and scalar multiplication as V).

Many of the ten vector space axioms are automatically satisfied since V is a vector space; indeed, to conclude that $H \subseteq V$ is a subspace of V, it suffices to check the following.

(a) The zero vector of V is in H.[2]

(b) H is closed under addition.

[1] One may think of the "field of scalars" as whatever set from which the "entries" the vectors come. For now, it is safe to assume $\mathbb{F} = \mathbb{R}$, but there is not much difference if $\mathbb{F} = \mathbb{C}$ or even a finite field.

[2] Equivalently, one could require that H be nonempty. Try to convince yourself that this condition is equivalent.

(c) H is closed under scalar multiplication.

Example 3.2.2. Let V be a vector space with $\mathbf{v}_1, \dots, \mathbf{v}_p \in V$. Then

$$\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \coloneqq \{c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p : c_1, \dots, c_p \in \mathbb{F}\}$$

is a subspace of V. In particular, this means that if we are able to express some set of vectors as a span, then that set is indeed a subspace.

Exercise 3.2.3. Show that the set

$$\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : a + 3b = c,\ b + c + a = d \right\}$$

is a subspace of $\mathbb{R}^4$ be writing it as a span.

Example 3.2.4. Let A be an $m \times n$ matrix. We define the **null space** of A and the **column space** of A to be the sets

$$\operatorname{Nul} A \coloneqq \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}, \quad \operatorname{Col} A \coloneqq \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}.$$

It can then be shown that $\operatorname{Nul} A$ is subspace of $\mathbb{R}^n$, and likewise $\operatorname{Col} A$ is a subspace of $\mathbb{R}^m$.

3.3 Linear Transformations

Definition 3.3.1. Let V and W be vector spaces. A transformation $T : V \to W$ is **linear** if

(a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$,

(b) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u} \in V$ and $c \in \mathbb{F}$.

Example 3.3.2. The transformation $D : \mathbb{P}_2 \to \mathbb{P}_1$ given by differentiation is linear.

Definition 3.3.3. Let $T : V \to W$ be a linear transformation. The **kernel** of T is the set

$$\{\mathbf{u} \in V : T(\mathbf{u}) = \mathbf{0}\}$$

and likewise the **range** of T is the set

$$\{T(\mathbf{x}) \in W : \mathbf{x} \in V\}.$$

Exercise 3.3.4. Define a linear transformation $T : \mathbb{P}_2 \to \mathbb{R}^2$ by

$$T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(0) \end{bmatrix}.$$

Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T, and describe the range of T.

3.4 Bases & Dimension

Definition 3.4.1. Let V be a vector space. We say that a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \subseteq V$ is **linearly independent** if the equation

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution $c_1 = \cdots = c_p = 0$. Otherwise, the set is **linearly dependent**.

Definition 3.4.2. Let V be a vector space with subspace H. An indexed set of vectors $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\} \subseteq V$ is a **basis** for H if $\mathcal{B}$ is a linearly independent set such that $H = \operatorname{span} \mathcal{B}$.

Example 3.4.3. Many vector spaces have "standard" bases which should be treated as the default bases for these spaces unless a different basis is specified. For example, $\mathbb{R}^3$ has the standard basis

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\},$$

while the standard basis for $\mathbb{P}_2$ is the basis of monomials $\{1, t, t^2\}$. It should not be difficult to imagine the generalizing these bases for $\mathbb{R}^n$ and $\mathbb{P}_n$ for other values of n.

Exercise 3.4.4. Assuming that A is row equivalent to B, find bases for $\operatorname{Nul} A$ and $\operatorname{Col} A$.

$$A = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Proposition 3.4.5. If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

As a result, we call the number of vectors in a basis for V the **dimension** of V, denoted $\dim V$. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

Example 3.4.6. From the example above, we see that $\dim \mathbb{R}^3 = \dim \mathbb{P}_2 = 3$, since both have bases consisting of 3 vectors. Indeed, in general we have that

$$\dim \mathbb{R}^n = n, \quad \dim \mathbb{P}_n = n + 1.$$

Definition 3.4.7. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for a vector space V, and $\mathbf{x} \in V$. The **$\mathcal{B}$-coordinate vector** of $\mathbf{x}$ is the vector

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \text{ such that } c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n = \mathbf{x}.$$

Exercise 3.4.8. The set $\mathcal{B} = \{1 + t^2, t + t^2, 1 + 2t + t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 1 + 4t + 7t^2$ relative to $\mathcal{B}$.

3.5 Change of Basis

Proposition 3.5.1. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $P_{\mathcal{C} \leftarrow \mathcal{B}}$ such that

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

In particular, the columns of $P_{\mathcal{C} \leftarrow \mathcal{B}}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$:

$$P_{\mathcal{C} \leftarrow \mathcal{B}} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \ldots & [\mathbf{b}_n]_{\mathcal{C}} \end{bmatrix}.$$

Example 3.5.2. In practice, the change of basis matrix is computed via row reduction:

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n & | & \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{bmatrix} \sim \begin{bmatrix} I & | & P_{\mathcal{C} \leftarrow \mathcal{B}} \end{bmatrix}.$$

For example, the sets

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \end{bmatrix} \right\} \text{ and } \mathcal{C} = \left\{ \begin{bmatrix} -7 \\ 9 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \end{bmatrix} \right\}$$

are each bases for $\mathbb{R}^2$. To find the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$, we find by row reduction

$$\begin{bmatrix} -7 & -5 & 1 & -2 \\ 9 & 7 & -3 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 & -3/2 \\ 0 & 1 & -3 & 5/2 \end{bmatrix}$$

and so the change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is

$$P_{\mathcal{C} \leftarrow \mathcal{B}} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix}$$

Proposition 3.5.3. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ be bases of a vector space V. Then the change of basis matrices from $\mathcal{B}$ to $\mathcal{C}$ and from $\mathcal{C}$ to $\mathcal{B}$ are related in the following way:

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

4 Eigenthings

4.1 Eigenvalues & Eigenvectors

Definition 4.1.1. Let A be an $n \times n$ matrix. We say that a nonzero vector $\mathbf{x}$ is an **eigenvector** of A if $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, in which case λ is an **eigenvalue** of A.

Exercise 4.1.2. Verify that $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is an eigenvector of $\begin{bmatrix} 2 & 6 & 7 \\ 3 & 2 & 7 \\ 5 & 6 & 4 \end{bmatrix}$. What is its corresponding eigenvalue?

Remark 4.1.3. Given a matrix A and a candidate eigenvector $\mathbf{x}$, it is easy to check whether $\mathbf{x}$ is indeed an eigenvector of A; simply compute the product $A\mathbf{x}$ and check whether it is a scalar multiple of $\mathbf{x}$.

The situation is a bit harder if you are instead given a matrix A and a candidate eigenvalue λ, but the following allows us to deal with this case.

Proposition 4.1.4. Let A be an $n \times n$ matrix. Then a scalar λ is an eigenvalue of A if and only if the equation

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

has a nontrivial solution. If it does, such a nontrivial solution is an eigenvector corresponding to λ.

Proof. By definition, λ is an eigenvalue of A if $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero vector $\mathbf{x}$. Subtracting $\lambda\mathbf{x}$ from both sides, this is equivalent to saying that

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

for some nonzero vector $\mathbf{x}$. □

Corollary 4.1.5. By the invertible matrix theorem together with the previous proposition, λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. We shall call $\det(A - \lambda I)$ the **characteristic polyomial** of A. Hence, the eigenvalues of A are the roots of its characteristic polynomial.

Example 4.1.6. Determine whether $\lambda = 3$ is an eigenvalue of the matrix

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

If so, find a corresponding eigenvector.

Example 4.1.7. Find all eigenvalues for each of the matrices below

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -3 \end{bmatrix}.$$

4.2 Diagonalization

Remark 4.2.1. Suppose we are interested in computing A^{23} for some $n \times n$ matrix A, or in general, A^k for some large value of k. A naïve method would be to perform 22 matrix multiplications, but this is not ideal.

A slightly better approach is to write k as a sum of powers of 2, in this case $23 = 16 + 4 + 2 + 1$. We already know A, and can compute A^2 by squaring it. Once A^2 is known, we can square that to get A^4, then again for A^8, and finally A^{16}. We can then compute

$$A^{23} = A^{16} A^4 A^2 A,$$

requiring 7 matrix multiplications in total.

However, it turns out this is still not ideal. Indeed, if we can "diagonalize" A, then we can compute A^k for any power k with only 2 matrix multiplications. This method will leverage the fact that powers of diagonal matrices are easy to compute:

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \implies D^2 = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$$

Definition 4.2.2. Let A be an $n \times n$ matrix. We say that A is **diagonalizable** if it is similar to a diagonal matrix; that is, if we can write

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

for some invertible matrix P and some diagonal matrix D.

Remark 4.2.3. It is not hard to see why writing A in the above form could be useful. In particular,

$$A^k = (PDP^{-1})^k = \underbrace{PDP^{-1}PDP^{-1} \ldots PDP^{-1}}_{k \text{ times}} = PD^kP^{-1}$$

and since D is a diagonal matrix, D^k is easy to compute. Hence, given a diagonalization for A, computing A^k can be done very cheaply. However, not all $n \times n$ matrices A can be diagonalized.

Exercise 4.2.4. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.

Proposition 4.2.5. An $n \times n$ matrix A is diagonalizable if and only if A has n linear independent eigenvectors.

If it does, $A = PDP^{-1}$ where the columns of P are the n linearly independent eigenvectors of A, and the diagonal entries of D are the corresponding eigenvalues of A.

Corollary 4.2.6. An $n \times n$ matrix with n distinct eigenvalues is diagonalizable. Note that this condition is sufficient, but not necessary.

Example 4.2.7. As a lower triangular matrix, it is clear that the below matrix has repeated eigenvalues; however, it is still diagonalizable.

$$A = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 1 & 4 & -3 & 0 \\ -1 & -2 & 0 & -3 \end{bmatrix}$$

4.3 Deriving Binet's Formula

Example 4.3.1. The Fibonacci sequence is defined by the linear recurrence relation

$$F_n = F_{n-1} + F_{n-2}, \quad F_0 = 1, \quad F_1 = 1.$$

It follows that the first few terms of the sequence are given by

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots.$$

We will now describe a method for deriving Binet's formula for F_n based on diagonalization; observe that we can rewrite the linear recurrence as the matrix equation

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2 \begin{bmatrix} F_{n-2} \\ F_{n-3} \end{bmatrix} = \cdots = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-1} \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}.$$

Hence, we would like to diagonalize the matrix

$$A \coloneqq \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

whose characteristic polynomial $\det(A - \lambda I) = \lambda^2 - \lambda - 1$ has the roots

$$\lambda = \frac{1 \pm \sqrt{5}}{2}$$

by the quadratic formula. We can then obtain that

$$v_1 = \begin{bmatrix} 1 + \sqrt{5} \\ 2 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 - \sqrt{5} \\ 2 \end{bmatrix}$$

are eigenvectors corresponding to $\lambda_1 = (1 + \sqrt{5})/2$ and $\lambda_2 = (1 - \sqrt{5})/2$ respectively, giving us the diagonalization

$$A = PDP^{-1} \text{ where } P = \begin{bmatrix} 1 + \sqrt{5} & 1 - \sqrt{5} \\ 2 & 2 \end{bmatrix} \text{ and } D = \frac{1}{2} \begin{bmatrix} 1 + \sqrt{5} & 0 \\ 0 & 1 - \sqrt{5} \end{bmatrix}.$$

In particular, then

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \frac{1}{20 \cdot 2^{n-1}} \begin{bmatrix} 1 + \sqrt{5} & 1 - \sqrt{5} \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 + \sqrt{5} & 0 \\ 0 & 1 - \sqrt{5} \end{bmatrix}^{n-1} \begin{bmatrix} 2\sqrt{5} & 5 - \sqrt{5} \\ -2\sqrt{5} & 5 + \sqrt{5} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

and so simplifying the right-hand side, using the fact that powers of diagonal matrices are taken element-wise, we obtain

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \frac{1}{2^n\sqrt{5}} \begin{bmatrix} (1+\sqrt{5})^n - (1-\sqrt{5})^n \\ 2(1+\sqrt{5})^{n-1} - 2(1-\sqrt{5})^{n-1} \end{bmatrix}.$$

Finally, observe that we recover Binet's formula from the first component:

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

Remark 4.3.2. The above process can easily be adapted to compute other constant-recursive sequences. For example, changing the initial values to 2 and 1 yields the Lucas sequence. Two other recurrences of interest are

$$s_n = 2s_{n-1} + s_{n-2}, \quad s_0 = 0, \quad s_1 = 1$$

which yields the Pell numbers, as well as

$$s_n = 3s_{n-1} - 3s_{n-2} + s_{n-3}, \quad s_0 = 1, \quad s_1 = 1, \quad s_2 = 3$$

which yields the triangular numbers. In the latter case, the matrix to be diagonalized will be 3×3.

5 Orthogonality

Remark 5.0.1. Recall that an inconsistent system is one which has no exact solutions. However, for some problems we may be interested in approximate solutions, which do not solve the system exactly but are as "close" to a solution as possible. In order to make precise this notion of "closeness" we will study the concepts of orthogonality.

5.1 Metrics, Norms, Inner Products, Oh My!

Definition 5.1.1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is the scalar quantity

$$\mathbf{u}^\top \mathbf{v} = \sum_{i=1}^n u_i v_i.$$

We will also denote this quantity as $\mathbf{u} \cdot \mathbf{v}$.

Proposition 5.1.2. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The following are immediate from the definition of inner product and arithmetic properties of real numbers.

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

(c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Definition 5.1.3. Let $\mathbf{v} \in \mathbb{R}^n$. The **norm** of $\mathbf{v}$ is the scalar quantity

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\sum_{i=1}^n v_i^2}.$$

If $\mathbf{v}$ is such that $\|\mathbf{v}\| = 1$, we say that $\mathbf{v}$ is a **unit vector**.

Proposition 5.1.4. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The following follow from properties of the inner product.

(a) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$

(b) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

(c) $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

Definition 5.1.5. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is the scalar quantity

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Properties analogous to those for the norm also follow for distance.

Exercise 5.1.6. Given the vectors

$$\mathbf{w} = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix},$$

compute the quantities $\mathbf{w} \cdot \mathbf{w}$, $\mathbf{x} \cdot \mathbf{w}$, $\frac{\mathbf{x} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}$, $\|w\|$, and $\|x\|$.

5.2 Orthogonality

Definition 5.2.1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$. More generally, a set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\} \subseteq \mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

Exercise 5.2.2. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}.$$

5.3 QR Factorization

5.4 Least Squares

Definition 5.4.1. Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. 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 $\mathbf{x} \in \mathbb{R}^n$.

Proposition 5.4.2. The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation, $A^\top A\mathbf{x} = A^\top \mathbf{b}$.

6 Symmetric Matrices

6.1 Spectral Theorem

Definition 6.1.1. A matrix A is said to be **symmetric** if $A^\top = A$. Observe that such a matrix is necessarily square.

Definition 6.1.2. Let A be an $n \times n$ matrix. We say that A is **orthogongally diagonalizable** if there exists an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^{\top}.$$

This is a special form of diagonalization where P is not simply invertible, but orthogonal, such that $P^{-1} = P^{\top}$.

Proposition 6.1.3. Let A be an $n \times n$ matrix. Then A is orthogonally diagonalizable if and only if A is a symmetric matrix.

Theorem 6.1.4 (Spectral Theorem)**.** Let A be an $n \times n$ symmetric matrix. Then 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) Eigenvectors corresponding to different eigenvalues are orthogonal.

(d) A is orthogonally diagonalizable.

Proposition 6.1.5 (Spectral Decomposition)**.** Let $A = PDP^{\top}$, with orthonormal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ and corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Then

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\top} + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^{\top} + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^{\top}.$$

7 Further Topics

7.1 Polar Decomposition

Theorem 7.1.1 (Polar Decomposition)**.** Let $A \in \mathcal{M}_n(\mathbb{F})$.

(a) There exist a unitary matrix U and a positive semi-definite matrix P such that $A = UP$.

(b) P is unique, and in particular, $P = \sqrt{A^* A}$.

(c) U is unique if A is invertible.

Remark 7.1.2. Recall that we can write complex numbers in polar form as $z = re^{i\theta}$. The polar decomposition can be thought of as an analog for matrices; U is playing a role analogous to $e^{i\theta}$, that is, the set of unitary matrices are analogous to the unit circle. Likewise, P is playing a role analogous to r, that is, the set of positive semi-definite matrices are analogous to the non-negative real numbers.

References

[1] D.C. Lay. *Linear Algebra and Its Applications.* Pearson Education, 2003.