Preface

TO BE WRITTEN.
# Contents

Notation ix

1 Matrix Theory 1

Introduction to Part I 2

1 \( (\mathbb{F}, \mathbb{R}) \) Column Vectors 3
   1.1 \( (\mathbb{F}) \) Column vector basics 3
      1.1.1 The domain of scalars 3
   1.2 \( (\mathbb{F}) \) Subspaces and span 6
   1.3 \( (\mathbb{F}) \) Linear independence and the First Miracle of Linear Algebra 8
   1.4 \( (\mathbb{F}) \) Dot product 12
   1.5 \( (\mathbb{R}) \) Dot product over \( \mathbb{R} \) 13
   1.6 \( (\mathbb{F}) \) Additional exercises 14

2 \( (\mathbb{F}) \) Matrices 15
   2.1 Matrix basics 15
   2.2 Matrix multiplication 18
   2.3 Arithmetic of diagonal and triangular matrices 22
   2.4 Permutation Matrices 23
   2.5 Additional exercises 25

3 \( (\mathbb{F}) \) Matrix Rank 28
   3.1 Column and row rank 28
## CONTENTS

### 9 (R) Orthogonal Matrices
- 9.1 Orthogonal matrices .............................................. 64
- 9.2 Orthogonal similarity ............................................ 64
- 9.3 Additional exercises ................................................ 65

### 10 (R) The Spectral Theorem
- 10.1 Statement of the Spectral Theorem ............................. 66
- 10.2 Applications of the Spectral Theorem .......................... 66

### 11 (F, R) Bilinear and Quadratic Forms
- 11.1 (F) Linear and bilinear forms .................................... 68
- 11.2 (F) Multivariate polynomials ..................................... 69
- 11.3 (R) Quadratic forms ................................................ 71
- 11.4 (F) Geometric algebra (optional) ............................... 73

### 12 (C) Complex Matrices
- 12.1 Complex numbers .................................................. 75
- 12.2 Hermitian dot product in C^n .................................... 75
- 12.3 Hermitian and unitary matrices .................................. 77
- 12.4 Normal matrices and unitary similarity .......................... 78
- 12.5 Additional exercises ................................................. 79

### 13 (C, R) Matrix Norms
- 13.1 (R) Operator norm .................................................. 80
- 13.2 (R) Frobenius norm .................................................. 81
- 13.3 (C) Complex Matrices .............................................. 81

## II Linear Algebra of Vector Spaces

### Introduction to Part II .................................................. 83

### 14 (F) Preliminaries
- 14.1 Modular arithmetic ................................................. 84
- 14.2 Abelian groups ...................................................... 86
- 14.3 Fields ................................................................. 89
14.4 Polynomials ................................................................. 89

15 (F) Vector Spaces: Basic Concepts .............................................. 97
  15.1 Vector spaces .............................................................. 97
  15.2 Subspaces and span ...................................................... 99
  15.3 Linear independence and bases ........................................ 100
  15.4 The First Miracle of Linear Algebra ................................... 104
  15.5 Direct sums ............................................................... 105

16 (F) Linear Maps ................................................................. 107
  16.1 Linear map basics ....................................................... 107
  16.2 Isomorphisms ............................................................. 108
  16.3 The Rank-Nullity Theorem ............................................. 109
  16.4 Linear transformations ................................................ 110
    16.4.1 Eigenvectors, eigenvalues, eigenspaces ......................... 111
    16.4.2 Invariant subspaces .............................................. 112
  16.5 Coordinatization ....................................................... 114
  16.6 Change of basis ....................................................... 116

17 (F) Block Matrices (optional) ............................................... 118
  17.1 Block matrix basics .................................................... 118
  17.2 Arithmetic of block-diagonal and block-triangular matrices .......... 119

18 (F) Minimal Polynomials of Matrices and Linear Transformations (optional) 121
  18.1 The minimal polynomial .............................................. 121
  18.2 Minimal polynomials of linear transformations ..................... 122

19 (R) Euclidean Spaces ....................................................... 125
  19.1 Inner products ......................................................... 125
  19.2 Gram-Schmidt orthogonalization .................................... 128
  19.3 Isometries and orthogonality ....................................... 129
  19.4 First proof of the Spectral Theorem ................................ 131

20 (C) Hermitian Spaces ....................................................... 134
  20.1 Hermitian spaces ...................................................... 134
  20.2 Hermitian transformations .......................................... 137
CONTENTS

20.3 Unitary transformations ................................................................. 138
20.4 Adjoint transformations in Hermitian spaces ...................................... 139
20.5 Normal transformations ................................................................. 139
20.6 The Complex Spectral Theorem for normal transformations ................. 140

21 (ℝ, ℂ) The Singular Value Decomposition ................................. 141
21.1 The Singular Value Decomposition ................................................ 141
21.2 Low-rank approximation ............................................................... 142

22 (ℝ) Finite Markov Chains ................................................................. 144
22.1 Stochastic matrices ................................................................. 144
22.2 Finite Markov Chains ................................................................. 144
22.3 Digraphs ................................................................. 146
22.4 Digraphs and Markov Chains ...................................................... 146
22.5 Finite Markov Chains and undirected graphs .................................. 146
22.6 Additional exercises ................................................................. 146

23 More Chapters ........................................................................ 148

24 Hints ......................................................................................... 149
24.1 Column Vectors ................................................................. 149
24.2 Matrices ........................................................................ 151
24.3 Matrix Rank ........................................................................ 152
24.4 Theory of Systems of Linear Equations I: Qualitative Theory ............... 152
24.5 Affine and Convex Combinations .................................................. 152
24.6 The Determinant ................................................................. 153
24.7 Theory of Systems of Linear Equations II: Cramer's Rule ...................... 153
24.8 Eigenvectors and Eigenvalues ....................................................... 153
24.9 Orthogonal Matrices ................................................................. 153
24.10 The Spectral Theorem ............................................................... 153
24.11 Bilinear and Quadratic Forms ...................................................... 153
24.12 Complex Matrices ................................................................. 154
24.13 Matrix Norms ................................................................. 154
24.14 Preliminaries ................................................................. 154
24.15 Vector Spaces ................................................................. 154
24.16 Linear Maps ................................................. 155
24.17 Block Matrices .............................................. 155
24.18 Minimal Polynomials of Matrices and Linear Transformations .................................. 155
24.19 Euclidean Spaces .......................................... 155
24.20 Hermitian Spaces .......................................... 156
24.21 The Singular Value Decomposition .......... 156
24.22 Finite Markov Chains ................................. 156

25 Solutions ......................................................... 157
25.1 Column Vectors .............................................. 157
25.2 Matrices ...................................................... 157
25.3 Matrix Rank .................................................. 157
25.4 Theory of Systems of Linear Equations I: Qualitative Theory ..................................... 157
25.5 Affine and Convex Combinations ................. 157
25.6 The Determinant ............................................. 157
25.7 Theory of Systems of Linear Equations II: Cramer’s Rule ........................................... 157
25.8 Eigenvectors and Eigenvalues .......................... 157
25.9 Orthogonal Matrices ....................................... 158
25.10 The Spectral Theorem ..................................... 158
25.11 Bilinear and Quadratic Forms ......................... 158
25.12 Complex Matrices ......................................... 159
25.13 Matrix Norms .............................................. 159
25.14 Preliminaries ............................................... 159
25.15 Vector Spaces .............................................. 159
25.16 Linear Maps ................................................. 159
25.17 Block Matrices .............................................. 159
25.18 Minimal Polynomials of Matrices and Linear Transformations .................................. 159
25.19 Euclidean Spaces .......................................... 159
25.20 Hermitian Spaces .......................................... 159
25.21 The Singular Value Decomposition .......... 159
25.22 Finite Markov Chains ................................. 159
# Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{C} )</td>
<td>The field of complex numbers</td>
</tr>
<tr>
<td>diag</td>
<td>Diagonal matrix</td>
</tr>
<tr>
<td>im(( \varphi ))</td>
<td>Image of ( \varphi )</td>
</tr>
<tr>
<td>( \mathbb{N} )</td>
<td>The set ( {0, 1, 2, \ldots } ) of natural numbers</td>
</tr>
<tr>
<td>([n])</td>
<td>The set ( {1, \ldots, n} )</td>
</tr>
<tr>
<td>( \mathbb{Q} )</td>
<td>The field of rational numbers</td>
</tr>
<tr>
<td>( \mathbb{R} )</td>
<td>The field of real numbers</td>
</tr>
<tr>
<td>( \mathbb{R}^\times )</td>
<td>The set of real numbers excluding 0</td>
</tr>
<tr>
<td>( \mathbb{Z} )</td>
<td>The set ( {\ldots, -2, -1, 0, 1, 2, \ldots } ) of integers</td>
</tr>
<tr>
<td>( \mathbb{Z}^+ )</td>
<td>The set ( {1, 2, 3, \ldots } ) of positive integers</td>
</tr>
<tr>
<td>( \oplus )</td>
<td>Direct sum</td>
</tr>
<tr>
<td>( \leq )</td>
<td>Subspace</td>
</tr>
<tr>
<td>( \emptyset )</td>
<td>The empty set</td>
</tr>
<tr>
<td>( \perp )</td>
<td>Orthogonal</td>
</tr>
</tbody>
</table>
Part I

Matrix Theory
Introduction to Part I

TO BE WRITTEN.
Chapter 1

(F, R) Column Vectors

1.1 (F) Column vector basics

We begin with a discussion of column vectors.

Definition 1.1.1 (Column vector). A column vector of height $k$ is a list of $k$ numbers arranged in a column, written as

$$\begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_k
\end{pmatrix}.$$

The $k$ numbers in the column are referred to as the entries of the column vector; we will normally use lower case Greek letters such as $\alpha$, $\beta$, and $\zeta$ to denote these numbers. We denote column vectors by bold letters such as $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{b}$, $\mathbf{e}$, $\mathbf{f}$, etc., so we may write

$$\mathbf{v} = \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_k
\end{pmatrix}.$$

1.1.1 The domain of scalars

In general, the entries of column vectors will be taken from a “field,” denoted by $\mathbb{F}$. We shall refer to the elements of the field $\mathbb{F}$ as “scalars,” and we will normally denote scalars by lowercase Greek letters. We will discuss fields in detail in Section 14.3. Informally, a field is a set endowed with the operations of addition and multiplication which obey the rules familiar from the arithmetic of real numbers (commutativity, associativity, inverses, etc.). Examples of fields are $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{F}_p$, where $\mathbb{Q}$ is the set of rational numbers, $\mathbb{R}$ is the set of real numbers, $\mathbb{C}$ is the set of complex numbers, and $\mathbb{F}_p$ is the “integers modulo $p$” for a prime $p$, so $\mathbb{F}_p$ is a finite field of “order” $p$ (order = number of elements).

The reader who is not comfortable with finite fields or with the abstract concept of fields may ignore the exercises related to them and always take $\mathbb{F}$ to be $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$. In fact, taking $\mathbb{F}$ to be $\mathbb{R}$ suffices for all sections except those specifically related to $\mathbb{C}$.

We shall also consider integral vectors, i.e., vectors whose entries are integers. $\mathbb{Z}$ denotes the set of integers and $\mathbb{Z}^k$ the set of integral vectors of height $k$, so $\mathbb{Z}^k \subseteq \mathbb{Q}^k \subseteq \mathbb{R}^k \subseteq \mathbb{C}^k$. 

Babai: Discover Linear Algebra.
© 2016 László Babai.

This chapter last updated August 29, 2016
Note that $\mathbb{Z}$ is not a field (division does not work within $\mathbb{Z}$). Our notation $\mathbb{F}$ for the domain of scalars always refers to a field and therefore does not include $\mathbb{Z}$.

**Notation 1.1.2.** Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. The expression $\sum_{i=1}^k \alpha_i$ denotes the sum $\alpha_1 + \cdots + \alpha_k$. More generally, for an index set $I = \{i_1, i_2, \ldots, i_\ell\}$, the expression $\sum_{i \in I} \alpha_i$ denotes the sum $\alpha_{i_1} + \cdots + \alpha_{i_\ell}$.

**Convention 1.1.3 (Empty sum).** The empty sum, denoted $\sum_{i=1}^0 \alpha_i$ or $\sum_{i \in \emptyset} \alpha_i$, evaluates to zero.

**Definition 1.1.4 (The space $\mathbb{F}^k$).** For a domain $\mathbb{F}$ of scalars, we define the space $\mathbb{F}^k$ of column vectors of height $k$ over $\mathbb{F}$ by

$$\mathbb{F}^k := \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} \mid \alpha_i \in \mathbb{F} \right\} \quad (1.1)$$

**Definition 1.1.5 (Zero vector).** The zero vector in $\mathbb{F}^k$ is the vector

$$0_k := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1.2)$$

**Exercise 1.1.9.** Verify that vector addition is commutative, i.e., for $\mathbf{v}, \mathbf{w} \in \mathbb{F}^k$, we have $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.\quad (1.5)

**Exercise 1.1.10.** Verify that vector addition is associative, i.e., for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{F}^k$, we have $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.\quad (1.6)

We often write $\mathbf{0}$ instead of $0_k$ when the height of the vector is clear from context.

**Definition 1.1.6 (All-ones vector).** The all-ones vector in $\mathbb{F}^k$ is the vector

$$1_k := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (1.3)$$

We sometimes write $1$ instead of $1_k$.

**Definition 1.1.7 (Addition of column vectors).** Addition of column vectors of the same height is defined elementwise, i.e.,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_k + \beta_k \end{pmatrix} \quad (1.4)$$

**Example 1.1.8.** Let $\mathbf{v} = \begin{pmatrix} 2 \\ 6 \\ -1 \end{pmatrix}$ and let $\mathbf{w} = \begin{pmatrix} 0 \\ -3 \\ 2 \end{pmatrix}$. Then $\mathbf{v} + \mathbf{w} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$.
Column vectors also carry with them the notion of “scaling” by an element of $\mathbb{F}$.

**Definition 1.1.11** (Multiplication of a column vector by a scalar). Let $\mathbf{v} \in \mathbb{F}^k$ and let $\lambda \in \mathbb{F}$. Then the vector $\lambda \mathbf{v}$ is the vector $\mathbf{v}$ after each entry has been scaled (multiplied) by a factor of $\lambda$, i.e.,

$$
\lambda \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_k
\end{pmatrix} = 
\begin{pmatrix}
\lambda \alpha_1 \\
\lambda \alpha_2 \\
\vdots \\
\lambda \alpha_k
\end{pmatrix}.
$$

(1.7)

**Example 1.1.12.** Let $\mathbf{v} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$. Then

$$3 \mathbf{v} = \begin{pmatrix} 6 \\ -3 \\ 3 \end{pmatrix}.$$

**Definition 1.1.13** (Linear combination). Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{F}^n$. Then a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a sum of the form

$$
\sum_{i=1}^{m} \alpha_i \mathbf{v}_i
$$

where $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$.

**Example 1.1.14.** The following is a linear combination.

$$
\begin{pmatrix} 2 \\ -3 \\ 6 \\ 1 \end{pmatrix} + 4 \begin{pmatrix} -1 \\ 2 \\ 0 \\ -3 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 6 \\ 2 \\ 10 \\ -8 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 1 \\ 7 \end{pmatrix}.
$$

**Numerical exercise 1.1.15.** The following linear combinations are of the form $\alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c}$. Evaluate in two ways, as $(\alpha \mathbf{a} + \beta \mathbf{b}) + \gamma \mathbf{c}$ and as $\alpha \mathbf{a} + (\beta \mathbf{b} + \gamma \mathbf{c})$. Self-check: you must get the same answer.

(a) $\begin{pmatrix} \frac{3}{4} \\ 3 \\ -1 \end{pmatrix} - 2 \begin{pmatrix} -7 \\ 2 \end{pmatrix}$

(b) $-2 \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -4 \\ 6 \end{pmatrix} + \begin{pmatrix} 1 \\ 4 \end{pmatrix}$

(c) $-\begin{pmatrix} 3 \\ 7 \\ -1 \end{pmatrix} - 4 \begin{pmatrix} -2 \\ 3 \end{pmatrix} + \begin{pmatrix} 7 \\ 3 \end{pmatrix}$

**Exercise 1.1.16.** Express the vector $\begin{pmatrix} -5 \\ -1 \\ -2 \\ 11 \end{pmatrix}$ as a linear combination of the vectors $\begin{pmatrix} -2 \\ 1 \\ 7 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$.

**Exercise 1.1.17.**

(a) Express $\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ as a linear combination of the vectors $\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$.

First describe the nature of the problem you need to solve.
CHAPTER 1. \((\mathbb{F}, \mathbb{R})\) COLUMN VECTORS

(b) Give an “aha” proof that \[
\begin{pmatrix}
3 \\
1 \\
-2
\end{pmatrix}
cannot
be expressed as a linear combination of
\[
\begin{pmatrix}
-3 \\
1 \\
2
\end{pmatrix},
\begin{pmatrix}
2 \\
-2 \\
0
\end{pmatrix},
\begin{pmatrix}
3 \\
-8 \\
5
\end{pmatrix}
\]
An “aha” proof may not be easy to find but it has to be immediately convincing. This problem will be generalized in Ex. 1.2.7.

Exercise 1.1.18. To what does a linear combination of the empty list of vectors evaluate? (R Convention 1.1.3)

Let us now consider the system of linear equations
\[
\begin{align*}
\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= \beta_1 \\
\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= \beta_2 \\
&\vdots \\
\alpha_{k1}x_1 + \alpha_{k2}x_2 + \cdots + \alpha_{kn}x_n &= \beta_k
\end{align*}
\]
(1.8)
Given the \(\alpha_{ij}\) and the \(\beta_i\), we need to find \(x_1, \ldots, x_n\) that satisfy these equations. This is arguably one of the most fundamental problems of applied mathematics. We can rephrase this problem in terms of the vectors \(a_1, \ldots, a_n, b\), where
\[
a_j := \begin{pmatrix}
\alpha_{1j} \\
\alpha_{2j} \\
\vdots \\
\alpha_{kj}
\end{pmatrix}
\]
is the column of coefficients of \(x_j\) and the vector
\[
b := \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_k
\end{pmatrix}
\]
represents the right-hand side. With this notation, our system of linear equations takes the more concise form
\[
x_1a_1 + x_2a_2 + \cdots + x_na_n = b .
\]
(1.9)
The problem of solving the system of equations (1.8) therefore is equivalent to expressing the vector \(b\) as a linear combination of the \(a_i\).

1.2 \((\mathbb{F})\) Subspaces and span

Definition 1.2.1 (Subspace). A set \(W \subseteq \mathbb{F}^n\) is a subspace of \(\mathbb{F}^n\) (denoted \(W \leq \mathbb{F}^n\)) if it is closed under linear combinations.

Exercise 1.2.2. Let \(W \leq \mathbb{F}^k\). Show that \(0 \in W\). (Why is the empty set not a subspace?)

Proposition 1.2.3. \(W \leq \mathbb{F}^n\) if and only if
\[
\begin{align*}
&\text{(a) } 0 \in W \\
&\text{(b) If } u, v \in W, \text{ then } u + v \in W \\
&\text{(c) If } v \in W \text{ and } \alpha \in \mathbb{F}, \text{ then } \alpha v \in W.
\end{align*}
\]

Exercise 1.2.4. Show that, if \(W\) is a nonempty subset of \(\mathbb{F}^n\), then \((c)\) implies \((a)\)
1.2. (F) SUBSPACES AND SPAN

Exercise 1.2.5. Show

(a) \( \{0\} \leq F^k \);
(b) \( F^k \leq F^k \).

We refer to these as the trivial subspaces of \( F^k \).

Exercise 1.2.6.

(a) The set
\[
\left\{ \begin{pmatrix} \alpha_1 \\ 0 \\ \alpha_3 \end{pmatrix} \mid \alpha_1 = 2\alpha_3 \right\}
\]
is a subspace of \( \mathbb{R}^3 \).
(b) The set
\[
\left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \mid \alpha_2 = \alpha_1 + 7 \right\}
\]
is not a subspace of \( \mathbb{R}^3 \).

Exercise 1.2.7. Let
\[
W_k = \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} \in F^k \mid \sum_{i=1}^k \alpha_i = 0 \right\}
\]
Show that \( W_k \leq F^k \). This is the 0-weight subspace of \( F^k \).

Exercise 1.2.8. Prove that, for \( n \geq 2 \), the space \( \mathbb{R}^n \) has infinitely many subspaces.

Proposition 1.2.9. Let \( W_1, W_2 \leq F^n \). Then

(a) \( W_1 \cap W_2 \leq F^n \)
(b) The intersection of any (finite or infinite) collection of subspaces of \( F^n \) is also a subspace of \( F^n \).
(c) \( W_1 \cup W_2 \leq F^n \) if and only if \( W_1 \subseteq W_2 \) or \( W_2 \subseteq W_1 \).

Definition 1.2.10 (Span). Let \( S \subseteq F^n \). Then the span of \( S \), denoted span(\( S \)), is the smallest subspace of \( F^k \) containing \( S \), i.e.,

(a) \( \text{span} \ S \supseteq S \);
(b) \( \text{span} \ S \) is a subspace of \( F^k \);
(c) for every subspace \( W \leq F^n \), if \( S \subseteq W \) then \( \text{span} \ S \leq W \).

Fact 1.2.11. \( \text{span}(\emptyset) = \{0\} \). (Why?)

Theorem 1.2.12. Let \( S \subseteq F^n \). Then

(a) \( \text{span} \ S \) exists and is unique;
(b) \( \text{span}(S) = \bigcap_{S \subseteq W \leq F^n} W \). This is the intersection of all subspaces containing \( S \).

This theorem tells us that the span exists. The next theorem constructs all the elements of the span.

Theorem 1.2.13. For \( S \subseteq F^n \), span \( S \) is the set of all linear combinations of the finite subsets of \( S \). (Note that this is true even when \( S \) is empty. Why?)

Proposition 1.2.14. Let \( S \subseteq F^n \). Then \( S \leq F^n \) if and only if \( S = \text{span}(S) \).

Let \( S \subseteq F^n \).
Proposition 1.2.15. Prove that \( \text{span}(\text{span}(S)) = \text{span}(S) \). Prove this

(a) based on the definition;

(b) (linear combinations of linear combinations) based on Theorem 1.2.13.

Proposition 1.2.16 (Transitivity of span). Suppose \( R \subseteq \text{span}(T) \) and \( T \subseteq \text{span}(S) \). Then \( R \subseteq \text{span}(S) \).

Definition 1.2.17 (Sum of sets). Let \( A, B \subseteq \mathbb{F}^n \). Then \( A + B \) is the set

\[
A + B = \{ a + b \mid a \in A, b \in B \}.
\] (1.10)

Proposition 1.2.18. Let \( U_1, U_2 \leq \mathbb{F}^n \). Then \( U_1 + U_2 = \text{span}(U_1 \cup U_2) \).

1.3 (\( \mathbb{F} \)) Linear independence and the First Miracle of Linear Algebra

In this section, \( v_1, v_2, \ldots \) will denote column vectors of height \( k \), i.e., \( v_i \in \mathbb{F}^k \).

Definition 1.3.1 (List). A list of objects \( a_i \) is a function whose domain is an “index set” \( I \); we write the list as \( \{a_i \mid i \in I\} \). Most often our index set will be \( I = \{1, \ldots, n\} \). In this case, we write the list as \( (a_1, \ldots, a_n) \). The size of the list is \( |I| \).

Notation 1.3.2 (Concatenation of lists). Let \( L = (v_1, v_2, \ldots, v_\ell) \) and \( M = (w_1, w_2, \ldots, w_n) \) be lists. We denote by \( (L, M) \) the list obtained by concatenation \( L \) and \( M \), i.e., the list \( (v_1, v_2, \ldots, v_\ell, w_1, w_2, \ldots, w_n) \). If the list \( M \) has only one element, we omit the parentheses around the list, that is, we write \( (L, w) \) rather than \( (L, (w)) \).

Definition 1.3.3 (Trivial linear combination). The trivial linear combination of the vectors \( v_1, \ldots, v_n \) is the linear combination (Def. 1.1.13) \( 0v_1 + \cdots + 0v_n \) (all coefficients are 0).

Fact 1.3.4. The trivial linear combination evaluates to zero.

Definition 1.3.5 (Linear independence). The list \( (v_1, \ldots, v_n) \) is said to be linearly independent if the only linear combination that evaluates to zero is the trivial linear combination.

The list \( (v_1, \ldots, v_n) \) is linearly dependent if it is not linearly independent, i.e., if there exist scalars \( \alpha_1, \ldots, \alpha_n \), not all zero, such that

\[
\sum_{i=1}^{n} \alpha_i v_i = 0.
\]

Definition 1.3.6. If a list \( (v_1, \ldots, v_n) \) of vectors is linearly independent (dependent), we say that the vectors \( v_1, \ldots, v_n \) are linearly independent (dependent).

Definition 1.3.7. We say that a set of vectors is linearly independent if a list formed by its elements (in any order and without repetitions) is linearly independent.
1.3. (F) LINEAR INDEPENDENCE AND THE FIRST MIRACLE OF LINEAR ALGEBRA

Example 1.3.8. Let
\[ \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} -4 \\ 1 \\ 2 \end{pmatrix}. \]
Then \( \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \) are linearly independent vectors. It follows that the set \( \{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_2 \} \) is linearly independent while the list \( (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_2) \) is linearly dependent.

Note that the list \( (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_2) \) has four elements, but the set \( \{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_2 \} \) has three elements.

Exercise 1.3.9. Show that the vectors
\[ \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ -8 \\ 3 \end{pmatrix} \]
are linearly dependent.

Definition 1.3.10. We say that the vector \( \mathbf{w} \) depends on the list \( (\mathbf{v}_1, \ldots, \mathbf{v}_k) \) of vectors if \( \mathbf{w} \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \), i.e., if \( \mathbf{w} \) can be expressed as a linear combination of the \( \mathbf{v}_i \).

Proposition 1.3.11. The vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \) are linearly dependent if and only if there is some \( i \) such that \( \mathbf{v}_i \) depends on the other vectors in the list.

Exercise 1.3.12. Show that the list \( (\mathbf{v}, \mathbf{w}, \mathbf{v} + \mathbf{w}) \) is linearly dependent.

Exercise 1.3.13. Is the empty list linearly independent?

Exercise 1.3.14. Show that if a list is linearly independent, then any permutation of it is linearly independent.

Proposition 1.3.15. The list \( (\mathbf{v}, \mathbf{v}) \), consisting of the vector \( \mathbf{v} \in \mathbb{F}^n \) listed twice, is linearly dependent.

Exercise 1.3.16. Is there a vector \( \mathbf{v} \) such that the list \( (\mathbf{v}) \), consisting of a single item, is linearly dependent?

Exercise 1.3.17. Which vectors depend on the empty list?

Definition 1.3.18 (Sublist). A sublist of a list \( L \) is a list \( M \) consisting of some of the elements of \( L \), in the same order in which they appear in \( L \).

Examples 1.3.19. Let \( L = (a, b, c, b, d, e) \).

(a) The empty list is a sublist of \( L \).

(b) \( L \) is a sublist of itself.

(c) The list \( L_1 = (a, b, d) \) is a sublist of \( L \).

(d) The list \( L_2 = (b, b) \) is a sublist of \( L \).

(e) The list \( L_3 = (b, b, b) \) is not a sublist of \( L \).

(f) The list \( L_4 = (a, d, c, e) \) is not a sublist of \( L \).

Fact 1.3.20. Every sublist of a linearly independent list of vectors is linearly independent.

The following lemma is central to the proof of the First Miracle of Linear Algebra (Theorem 1.3.40) as well as to our characterization of bases as maximal linearly independent sets (Prop. 1.3.37).
Lemma 1.3.21. Suppose \((v_1, \ldots, v_k)\) is a linearly independent list of vectors and the list \((v_1, \ldots, v_{k+1})\) is linearly dependent. Then \(v_{k+1} \in \text{span}(v_1, \ldots, v_k)\). ♦

Proposition 1.3.22. The vectors \(v_1, \ldots, v_k\) are linearly dependent if and only if there is some \(j\) such that
\[
v_j \in \text{span}(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k).
\]

Exercise 1.3.23. Prove that no list of vectors containing \(0\) is linearly independent.

Exercise 1.3.24. Prove that a list of vectors with repeated elements (the same vector occurs more than once) is linearly dependent. (This follows from combining which two previous exercises?)

Definition 1.3.25 (Parallel vectors). Let \(u, v \in \mathbb{F}^n\). We say that \(u\) and \(v\) are parallel if there exists a scalar \(\alpha\) such that \(u = \alpha v\) or \(v = \alpha u\). Note that \(0\) is parallel to all vectors, and the relation of being parallel is an equivalence relation (\(\mathbb{F}^n\) Def. 14.1.19) on the set of nonzero vectors.

Exercise 1.3.26. Let \(u, v \in \mathbb{F}^n\). Show that the list \((u, v)\) is linearly dependent if and only if \(u\) and \(v\) are parallel.

Exercise 1.3.27. Find \(n+1\) vectors in \(\mathbb{F}^n\) such that every \(n\) are linearly independent. Over infinite fields, a much stronger statement holds (\(\mathbb{F}^n\) Ex. 15.3.15).

CHAPTER 1. \((\mathbb{F}, \mathbb{R})\) COLUMN VECTORS

Definition 1.3.28 (Rank). The rank of a set \(S \subseteq \mathbb{F}^n\), denoted \(\text{rk} S\), is the size of the largest linearly independent subset of \(S\). The rank of a list is the rank of the set formed by its elements.

Proposition 1.3.29. Let \(S\) and \(T\) be lists. Show that
\[
\text{rk}(S, T) \leq \text{rk} S + \text{rk} T \quad (1.11)
\]
where \(\text{rk}(S, T)\) is the rank of the list obtained by concatenating the lists \(S\) and \(T\).

Definition 1.3.30 (Dimension). Let \(W \subseteq \mathbb{F}^n\) be a subspace. The dimension of \(W\), denoted \(\dim W\), is its rank, that is, \(\dim W = \text{rk} W\).

Definition 1.3.31 (List of generators). Let \(W \subseteq \mathbb{F}^n\). The list \(L = (v_1, \ldots, v_k) \subseteq W\) is said to be a list of generators of \(W\) if \(\text{span}(v_1, \ldots, v_k) = W\). In this case, we say that \(v_1, \ldots, v_k\) generate \(W\).

Definition 1.3.32 (Basis). A list \(b = (b_1, \ldots, b_n)\) is a basis of the subspace \(W \subseteq \mathbb{F}^k\) if \(b\) is a linearly independent list generators of \(W\).

Fact 1.3.33. If \(W \subseteq \mathbb{F}^n\) and \(\tilde{b}\) is a list of vectors of \(W\) then \(\tilde{b}\) is a basis of \(W\) if and only if it is linearly independent and \(\text{span}(\tilde{b}) = W\).

Definition 1.3.34 (Standard basis of \(\mathbb{F}^k\)). The standard basis of \(\mathbb{F}^k\) is the basis \((e_1, \ldots, e_k)\), where \(e_i\) is the column vector which has its \(i\)-th component equal to 1 and all other components equal to 0. The vectors \(e_1, \ldots, e_k\) are sometimes called the standard unit vectors.
For example, the standard basis of \( \mathbb{F}^3 \) is
\[
\begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix}, \quad \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix}, \quad \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix}
\]

**Exercise 1.3.35.** The standard basis of \( \mathbb{F}^k \) is a basis.

Note that subspaces of \( \mathbb{F}^k \) do not have a “standard basis.”

**Definition 1.3.36 (Maximal linearly independent list).** Let \( W \leq \mathbb{F}^n \). A list \( L = (v_1, \ldots, v_k) \) of vectors in \( W \) is a maximal linearly independent list in \( W \) if \( L \) is linearly independent but for all \( w \in W \), the list \((L, w)\) is linearly dependent.

**Proposition 1.3.37.** Let \( W \leq \mathbb{F}^k \) and let \( b = (b_1, \ldots, b_k) \) be a list of vectors in \( W \). Then \( b \) is a basis of \( W \) if and only if it is a maximal linearly independent list.

**Exercise 1.3.38.** Prove: every subspace \( W \leq \mathbb{F}^k \) has a basis. You may assume the fact, to be proved shortly (Cor. 1.3.49), that every linearly independent list of vectors in \( \mathbb{F}^k \) has size at most \( k \).

**Proposition 1.3.39.** Let \( L \) be a finite list of generators of \( W \leq \mathbb{F}^k \). Then \( L \) contains a basis of \( W \).

Next we state a central result to which the entire field of linear algebra arguably owes its character.

**Theorem 1.3.40 (First Miracle of Linear Algebra).** Let \( v_1, \ldots, v_k \) be linearly independent with \( v_i \in \text{span}(w_1, \ldots, w_m) \) for all \( i \). Then \( k \leq m \).

The proof of this theorem requires the following lemma.

**Lemma 1.3.41 (Steinitz exchange lemma).** Let \( (v_1, \ldots, v_k) \) be a linearly independent list of vectors such that \( v_i \in \text{span}(w_1, \ldots, w_m) \) for all \( i \). Then there exists \( j \) (1 \( \leq j \leq m \)) such that the list \((w_j, v_2, \ldots, v_k)\) is linearly independent.

**Exercise 1.3.42.** Use the Steinitz exchange lemma to prove the First Miracle of Linear Algebra.

**Corollary 1.3.43.** Let \( W \leq \mathbb{F}^k \). Every basis of \( W \) has the same size (same number of vectors).

**Exercise 1.3.44.** Prove: Cor. 1.3.43 is equivalent to the First Miracle, i.e., infer the First Miracle from Cor. 1.3.43.

**Corollary 1.3.45.** Every basis of \( \mathbb{F}^k \) has size \( k \).

The following result is essentially a restatement of the First Miracle of Linear Algebra.

**Corollary 1.3.46.** \( \text{rk}(v_1, \ldots, v_k) = \dim(\text{span}(v_1, \ldots, v_k)) \).

**Exercise 1.3.47.** Prove Cor. 1.3.46 is equivalent to the First Miracle, i.e., infer the First Miracle from Cor. 1.3.46.

**Corollary 1.3.48.** \( \dim \mathbb{F}^k = k \).
Corollary 1.3.49. Every linearly independent list of vectors in $\mathbb{F}^k$ has size at most $k$.

Corollary 1.3.50. Let $W \leq \mathbb{F}^k$ and let $L$ be a linearly independent list of vectors in $W$. Then $L$ can be extended to a basis of $W$.

Exercise 1.3.51. Let $U_1, U_2 \leq \mathbb{F}^n$ with $U_1 \cap U_2 = \{0\}$. Let $v_1, \ldots, v_k \in U_1$ and $w_1, \ldots, w_\ell \in U_2$. If the lists $(v_1, \ldots, v_k)$ and $(w_1, \ldots, w_\ell)$ are linearly independent, then so is the concatenated list $(v_1, \ldots, v_k, w_1, \ldots, w_\ell)$.

Proposition 1.3.52. Let $U_1, U_2 \leq \mathbb{F}^n$ with $U_1 \cap U_2 = \{0\}$. Then
\[ \dim U_1 + \dim U_2 \leq n. \quad (1.12) \]

Proposition 1.3.53 (Modular equation). Let $U_1, U_2 \leq \mathbb{F}^k$. Then
\[ \dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2. \quad (1.13) \]

1.4 $\mathbb{F}$ Dot product

We now define the dot product, an operation which takes two vectors as input and outputs a scalar.

Definition 1.4.1 (Dot product). Let $x, y \in \mathbb{F}^n$ with $x = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$ and $y = (\beta_1, \beta_2, \ldots, \beta_n)^T$. (Note that these are column vectors). Then the dot product of $x$ and $y$, denoted $x \cdot y$, is
\[ x \cdot y := \sum_{i=1}^{n} \alpha_i \beta_i \in \mathbb{F}. \quad (1.14) \]

CHAPTER 1. $\mathbb{F}$, $\mathbb{R}$) COLUMN VECTORS

Proposition 1.4.2. The dot product is symmetric, that is,
\[ x \cdot y = y \cdot x. \]

Exercise 1.4.3. Show that the dot product is distributive, that is,
\[ x \cdot (y + z) = x \cdot y + x \cdot z. \]

Exercise 1.4.4. Show that the dot product is blinear ($\mathbb{F}$ Def. 6.2.11), i.e., for $x, y, z \in \mathbb{F}^n$ and $\alpha \in \mathbb{F}$, we have
\[ (x + z) \cdot y = x \cdot y + z \cdot y \quad (1.15) \]
\[ (\alpha x) \cdot y = \alpha (x \cdot y) \quad (1.16) \]
\[ x \cdot (y + z) = x \cdot y + x \cdot z \quad (1.17) \]
\[ x \cdot (\alpha y) = \alpha (x \cdot y) \quad (1.18) \]

Exercise 1.4.5. Show that the dot product preserves linear combinations, i.e., for $x_1, \ldots, x_k, y \in \mathbb{F}^n$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, we have
\[ \left( \sum_{i=1}^{k} \alpha_i x_i \right) \cdot y = \sum_{i=1}^{k} \alpha_i (x_i \cdot y) \quad (1.19) \]
and for $x, y_1, \ldots, y_\ell \in \mathbb{F}^n$ and $\beta_1, \ldots, \beta_\ell \in \mathbb{F}$,
\[ x \cdot \left( \sum_{i=1}^{\ell} \beta_i y_i \right) = \sum_{i=1}^{\ell} \beta_i (x \cdot y_i). \quad (1.20) \]

Numerical exercise 1.4.6. Let
\[ x = \begin{pmatrix} 3 \\ 4 \\ -2 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad z = \begin{pmatrix} 3 \\ 2 \\ -4 \end{pmatrix} \]

Compute $x \cdot y$, $x \cdot z$, and $x \cdot (y + z)$. Self-check: verify that $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.
Exercise 1.4.7. Compute the following dot products.

(a) $1_k \cdot 1_k$ for $k \geq 1$.

(b) $x \cdot x$ where $x = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} \in \mathbb{F}^k$

Definition 1.4.8 (Orthogonality). Let $x, y \in \mathbb{F}^k$. Then $x$ and $y$ are orthogonal (notation: $x \perp y$) if $x \cdot y = 0$.

Exercise 1.4.9. What vectors are orthogonal to every vector?

Exercise 1.4.10. Which vectors are orthogonal to $1_k$?

Definition 1.4.11 (Isotropic vector). The vector $v \in \mathbb{F}^n$ is isotropic if $v \neq 0$ and $v \cdot v = 0$.

Exercise 1.4.12.

(a) Show that there are no isotropic vectors in $\mathbb{R}^n$.

(b) Find an isotropic vector in $\mathbb{C}^2$.

(c) Find an isotropic vector in $\mathbb{F}_2^2$.

Theorem 1.4.13. If $v_1, \ldots, v_k$ are pairwise orthogonal and non-isotropic vectors, then they are linearly independent.

1.5 (\(\mathbb{R}\)) Dot product over \(\mathbb{R}\)

We now specialize our discussion of the dot product to the space $\mathbb{R}^n$.

Exercise 1.5.1. Let $x \in \mathbb{R}^n$. Show $x \cdot x \geq 0$, with equality holding if and only if $x = 0$.

This “positive definiteness” of the dot product allows us to define the “norm” of a vector, a generalization of length.

Definition 1.5.2 (Norm). The norm of a vector $x \in \mathbb{R}^n$, denoted $\|x\|$, is defined as

$$\|x\| := \sqrt{x \cdot x}.$$ (1.21)

This norm is also referred to as the Euclidean norm or the $\ell^2$ norm.

Definition 1.5.3 (Orthogonal system). An orthogonal system in $\mathbb{R}^n$ is a list of (pairwise) orthogonal nonzero vectors in $\mathbb{R}^n$.

Exercise 1.5.4. Let $S \subseteq \mathbb{R}^k$ be an orthogonal system in $\mathbb{R}^k$. Prove that $S$ is linearly independent.

Definition 1.5.5 (Orthonormal system). An orthonormal system in $\mathbb{R}^n$ is a list of (pairwise) orthogonal vectors in $\mathbb{V}$, all of which have unit norm.

Definition 1.5.6 (Orthonormal basis). An orthonormal basis of $W \leq \mathbb{R}^n$ is an orthonormal system that is a basis of $W$.

Exercise 1.5.7.

(a) Find an orthonormal basis of $\mathbb{R}^n$.  

(b) Find all orthonormal bases of $\mathbb{R}^2$.

Orthogonality is studied in more detail in Chapter 19.

1.6 ($\mathbb{F}$) Additional exercises

Exercise 1.6.1. Let $v_1, \ldots, v_n$ be vectors such that $\|v_i\| > 1$ for all $i$, and $v_i \cdot v_j = 1$ whenever $i \neq j$. Show that $v_1, \ldots, v_n$ are linearly independent.

Definition 1.6.2 (Incidence vector). Let $A \subseteq \{1, \ldots, n\}$. Then the incidence vector $v_A \in \mathbb{R}^n$ is the vector whose $i$-th coordinate is 1 if $i \in A$ and 0 otherwise.

Exercise 1.6.3. Let $A, B \subseteq \{1, \ldots, n\}$. Express the dot product $v_A \cdot v_B$ in terms of the sets $A$ and $B$.

♥ Exercise 1.6.4 (Generalized Fisher Inequality). Let $\lambda \geq 1$ and let $A_1, \ldots, A_m \subseteq \{1, \ldots, n\}$ such that for all $i \neq j$ we have $|A_i \cap A_j| = \lambda$. Then $m \leq n$. 
Chapter 2

(\mathbb{F}) Matrices

2.1 Matrix basics

Definition 2.1.1 (Matrix). A $k \times n$ matrix is a table of numbers arranged in $k$ rows and $n$ columns, written as

$$
\begin{pmatrix}
\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{k1} & \alpha_{k2} & \cdots & \alpha_{kn}
\end{pmatrix}
$$

We may write $M = (\alpha_{i,j})_{k \times n}$ to indicate a matrix whose entry in position $(i, j)$ ($i$-th row, $k$-th column) is $\alpha_{i,j}$. For typographical convenience we usually omit the comma separating the row index and the column index and simply write $\alpha_{ij}$ instead of $\alpha_{i,j}$; we use the comma if its omission would lead to ambiguity. So we write $M = (\alpha_{ij})_{k \times n}$, or simply $M = (\alpha_{ij})$ if the values $k$ and $n$ are clear from context. We also write $(M)_{ij}$ to indicate the $(i, j)$ entry of the matrix $M$, i.e., $\alpha_{ij} = (M)_{ij}$.

Example 2.1.2. This is a $3 \times 5$ matrix.

$$
\begin{pmatrix}
1 & 3 & -2 & 0 & -1 \\
6 & 5 & 4 & -3 & -6 \\
2 & 7 & 1 & 1 & 5
\end{pmatrix}
$$

In this example, we have $\alpha_{22} = 5$ and $\alpha_{15} = -1$.

Definition 2.1.3 (The space $\mathbb{F}^{k \times n}$). The set of $k \times n$ matrices with entries from the domain $\mathbb{F}$ of scalars is denoted by $\mathbb{F}^{k \times n}$. Recall that $\mathbb{F}$ always denotes a field. Square matrices ($k = n$) have special significance, so we write $M_n(\mathbb{F}) := \mathbb{F}^{n \times n}$. We identify $M_1(\mathbb{F})$ with $\mathbb{F}$ and omit the matrix notation, i.e., we write $\alpha$ rather than $(\alpha)$. An integral matrix is a matrix with integer entries. Naturally, $\mathbb{Z}^{k \times n}$ will denote the set of $k \times n$ integral matrices, and $M_n(\mathbb{Z}) = \mathbb{Z}^{n \times n}$. Recall that $\mathbb{Z}$ is not a field.

Example 2.1.4. For example, \( \begin{pmatrix} 0 & -1 & 4 & 7 \\ -3 & 5 & 6 & 8 \end{pmatrix} \in \mathbb{R}^{2 \times 4} \) is a $2 \times 4$ matrix and \( \begin{pmatrix} 2 & 6 & 9 \\ 3 & -4 & -2 \\ -5 & 1 & 4 \end{pmatrix} \in M_3(\mathbb{R}) \) is a $3 \times 3$ matrix.

Observe that the column vectors of height $k$ introduced in Chapter 1 are $k \times 1$ matrices, so $\mathbb{F}^k = \mathbb{F}^{k \times 1}$. Moreover, every statement about column vectors in Chapter 1 applies analogously to $1 \times n$ matrices (“row vectors”).

Notation 2.1.5. When writing row vectors, we use commas to avoid ambiguity, so we write, for example, $(3, 5, -1)$ instead of $(3 \ 5 \ -1)$. 

---

Babai: Discover Linear Algebra. 
© 2016 László Babai.
**CHAPTER 2. \((\mathbb{F})\) MATRICES**

**Notation 2.1.6 (Zero matrix).** The \(k \times n\) matrix with all of its entries equal to 0 is called the zero matrix and is denoted by \(0_{k \times n}\), or simply by 0 if \(k\) and \(n\) are clear from context.

**Notation 2.1.7 (All-ones matrix).** The \(k \times n\) matrix with all of its entries equal to 1 is denoted by \(J_{k \times n}\) or \(J\). We write \(J_n\) for \(J_{n \times n}\).

**Definition 2.1.8 (Diagonal matrix).** A matrix \(A = (\alpha_{ij}) \in M_n(\mathbb{F})\) is diagonal if \(\alpha_{ij} = 0\) whenever \(i \neq j\). The \(n \times n\) diagonal matrix with entries \(\lambda_1, \ldots, \lambda_n\) is denoted by \(\text{diag}(\lambda_1, \ldots, \lambda_n)\).

**Example 2.1.9.**

\[
\text{diag}(5, 3, 0, -1, 5) = \begin{pmatrix}
5 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 5
\end{pmatrix}
\]

**Notation 2.1.10.** To avoid filling most of a matrix with the number “0”, we often write matrices like the one above as

\[
\text{diag}(5, 3, 0, -1, 5) = \begin{pmatrix}
5 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 5
\end{pmatrix}
\]

where the big 0 symbol means that every entry in the triangles above or below the diagonal is 0.

**Definition 2.1.11 (Upper and lower triangular matrices).** A matrix \(A = (\alpha_{ij}) \in M_n(\mathbb{F})\) is upper triangular if \(\alpha_{ij} = 0\) whenever \(i > j\). A is said to be strictly upper triangular if \(\alpha_{ij} = 0\) whenever \(i \geq j\). Lower triangular and strictly lower triangular matrices are defined analogously.

**Examples 2.1.12.**

(a) \[
\begin{pmatrix}
5 & 2 & 0 & 7 & 2 \\
3 & 0 & -4 & 0 & 0 \\
0 & 6 & 0 & 0 & 0 \\
0 & -1 & -3 & 0 & 0 \\
0 & 0 & 0 & 0 & 5
\end{pmatrix}
\]

is upper triangular.

(b) \[
\begin{pmatrix}
0 & 2 & 0 & 7 & 2 \\
0 & 0 & -4 & 0 & 0 \\
0 & 6 & 0 & 0 & 0 \\
0 & 0 & -3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

is strictly upper triangular.

(c) \[
\begin{pmatrix}
5 & 0 & 2 & 3 \\
0 & 0 & 0 & 0 \\
7 & -4 & 6 & -1 \\
2 & 0 & 0 & 0
\end{pmatrix}
\]

is lower triangular.

(d) \[
\begin{pmatrix}
0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 \\
7 & -4 & 6 & 0 \\
2 & 0 & 0 & -3
\end{pmatrix}
\]

is strictly lower triangular.

**Fact 2.1.13.** The diagonal matrices are the matrices which are simultaneously upper and lower triangular.
2.1. MATRIX BASICS

Definition 2.1.14 (Matrix transpose). The transpose of a $k \times \ell$ matrix $M = (\alpha_{ij})$ is the $\ell \times k$ matrix $(\beta_{ij})$ defined by

$$\beta_{ij} = \alpha_{ji} \quad (2.1)$$

and is denoted $M^T$. (We flip it across its main diagonal, so the rows of $A$ become the columns of $A^T$ and vice versa.)

Examples 2.1.15.

(a) $\begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{pmatrix}^T = \begin{pmatrix} 3 & 1 \\ 1 & 5 \\ 4 & 9 \end{pmatrix}$

(b) $1_k = (1,1,\ldots,1)^T$ $k$ times

(c) In Examples 2.1.12, the matrix (c) is the transpose of (a) and (d) is the transpose of (b).

Fact 2.1.16. Let $A$ be a matrix. Then $(A^T)^T = A$.

Definition 2.1.17 (Symmetric matrix). A matrix $M$ is symmetric if $M = M^T$.

Note that if a matrix $M \in \mathbb{F}^{k \times \ell}$ is symmetric then $k = \ell$ ($M$ is square).

Example 2.1.18. The matrix $\begin{pmatrix} 1 & 3 & 0 \\ 3 & 5 & -2 \\ 0 & -2 & 4 \end{pmatrix}$ is symmetric.

Definition 2.1.19 (Matrix addition). Let $A = (\alpha_{ij})$ and $B = (\beta_{ij})$ be $k \times n$ matrices. Then the sum $A + B$ is the $k \times n$ matrix with entries

$$(A + B)_{ij} = \alpha_{ij} + \beta_{ij} \quad (2.2)$$

That is, addition is defined elementwise.

Example 2.1.20. $\begin{pmatrix} 2 & 1 \\ 4 & -2 \\ 0 & 5 \end{pmatrix} + \begin{pmatrix} 1 & -6 \\ 2 & 3 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 3 & -5 \\ 6 & 1 \\ 1 & 9 \end{pmatrix}$

Fact 2.1.21 (Adding zero). For any matrix $A \in \mathbb{F}^{k \times n}$, $A + 0 = 0 + A = A$.

Proposition 2.1.22 (Commutativity). Matrix addition obeys the commutative law: if $A,B \in \mathbb{F}^{k \times n}$, then $A + B = B + A$.

Proposition 2.1.23 (Associativity). Matrix addition obeys the associative law: if $A,B,C \in \mathbb{F}^{k \times n}$, then $(A + B) + C = A + (B + C)$.

Definition 2.1.24 (The negative of a matrix). Let $A \in \mathbb{F}^{k \times n}$ be a matrix. Then $-A$ is the $k \times n$ matrix defined by $(-A)_{ij} = -\alpha_{ij}$.

Proposition 2.1.25. Let $A \in \mathbb{F}^{k \times n}$. Then $A + (-A) = 0$.

Definition 2.1.26 (Multiplication of a matrix by a scalar). Let $A = (\alpha_{ij}) \in \mathbb{F}^{k \times n}$, and let $\zeta \in \mathbb{F}$. Then $\zeta A$ is the $k \times n$ matrix whose $(i,j)$ entry is $\zeta \cdot \alpha_{ij}$.

Example 2.1.27. $3 \begin{pmatrix} 1 & 2 \\ -3 & 4 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ -9 & 12 \\ 0 & 18 \end{pmatrix}$

Fact 2.1.28. $-A = (-1)A$. 


2.2 Matrix multiplication

Definition 2.2.1 (Matrix multiplication). Let $A = (\alpha_{ij})$ be an $r \times s$ matrix and $B = (\beta_{jk})$ be an $s \times t$ matrix. Then the matrix product $C = AB$ is the $r \times t$ matrix $C = (\gamma_{ik})$ defined by

$$\gamma_{ik} = \sum_{j=1}^{s} \alpha_{ij} \beta_{jk} \quad (2.3)$$

Exercise 2.2.2. Let $A \in \mathbb{F}^{k \times n}$ and let $B \in \mathbb{F}^{n \times m}$. Show that 

$$(AB)^T = B^T A^T . \quad (2.4)$$

Proposition 2.2.3 (Distributivity). Matrix multiplication obeys the right distributive law: if $A \in \mathbb{F}^{k \times n}$ and $B, C \in \mathbb{F}^{n \times \ell}$, then $A(B + C) = AB + AC$. Analogously, it obeys the left distributive law: if $A, B \in \mathbb{F}^{k \times n}$ and $C \in \mathbb{F}^{n \times \ell}$, then $(A + B)C = AC + BC$.

Proposition 2.2.4 (Associativity). Matrix multiplication obeys the associative law: if $A$, $B$, and $C$ are matrices with compatible dimensions, then $(AB)C = A(BC)$.

Proposition 2.2.5 (Matrix multiplication vs. scaling). Let $A \in \mathbb{F}^{k \times n}$, $B \in \mathbb{F}^{n \times \ell}$, and $\alpha \in \mathbb{F}$. Then

$$A(\alpha B) = \alpha(AB) = (\alpha A)B . \quad (2.5)$$

Numerical exercise 2.2.6. For each of the following triples of matrices, compute the products $AB$, $AC$, and $A(B+C)$. Self-check: verify that $A(B + C) = AB + AC$.

(a) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$,  
$$B = \begin{pmatrix} 3 & 1 & 0 \\ -4 & 2 & 5 \end{pmatrix} ,$$  
$$C = \begin{pmatrix} 1 & -7 & -4 \\ 5 & 3 & -6 \end{pmatrix}$$

(b) $A = \begin{pmatrix} 2 & 5 \\ 1 & 1 \\ 3 & -3 \end{pmatrix}$,  
$$B = \begin{pmatrix} 4 & 6 & 3 & 1 \\ 3 & 3 & -5 & 4 \end{pmatrix} ,$$  
$$C = \begin{pmatrix} 1 & -4 & -1 & 5 \\ 2 & 4 & 10 & -7 \end{pmatrix}$$

(c) $A = \begin{pmatrix} 4 & -2 & 2 \\ 1 & -3 & -2 \end{pmatrix}$,  
$$B = \begin{pmatrix} -1 & 4 \\ 3 & 2 \\ -5 & -2 \end{pmatrix} ,$$  
$$C = \begin{pmatrix} 2 & -3 \\ -6 & -2 \\ 0 & 1 \end{pmatrix}$$

Exercise 2.2.7. Let $A \in \mathbb{F}^{k \times n}$, and let $E_{ij}^{(\ell \times m)}$ be the $\ell \times m$ matrix with a 1 in the $(i, j)$ position and 0 everywhere else.

(a) What is $E_{ij}^{(k \times k)} A$?

(b) What is $AE_{ij}^{(n \times n)}$?

Definition 2.2.8 (Rotation matrix). The rotation matrix $R_\theta$ is the matrix defined by

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} . \quad (2.6)$$
As we shall see, this matrix is intimately related to the rotation of the Euclidean plane by \( \theta \) (Example 16.5.2).

**Exercise 2.2.9.** Prove \( R_{\alpha + \beta} = R_{\alpha}R_{\beta} \). Your proof may use the addition theorems for the trigonometric functions. Later when we learn about the connection between matrices and linear transformations, we shall give a direct proof of this fact which will imply the addition theorems (Example 16.5.12).

**Definition 2.2.10 (Identity matrix).** The \( n \times n \) identity matrix, denoted \( I_n \) or \( I \), is the diagonal matrix whose diagonal entries are all 1, i.e.,

\[
I = \text{diag}(1, 1, \ldots, 1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad (2.7)
\]

This is also written as \( I = (\delta_{ij}) \), where the Kronecker delta symbol \( \delta_{ij} \) is defined by

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases} \quad (2.8)
\]

**Fact 2.2.11.** The columns of \( I \) are the standard unit vectors (Example 1.3.34).

**Proposition 2.2.12.** For all \( A \in \mathbb{F}^{k \times n} \),

\[
I_k A = A I_n = A . \quad (2.9)
\]
(d) Interpret and verify the following statements:

(d1) $A^k$ grows linearly

(d2) $B^k$ grows exponentially

(d3) $C^k$ stays of constant “size.” (Define “size” in this statement.)

Definition 2.2.18 (Nilpotent matrix). The matrix $N \in M_n(\mathbb{F})$ is nilpotent if there exists an integer $k$ such that $N^k = 0$.

Exercise 2.2.19. Show that if $A \in M_n(\mathbb{F})$ is strictly upper triangular, then $A^n = 0$.

So every strictly upper triangular matrix is nilpotent. Later we shall see that a matrix is nilpotent if and only if it is “similar” (Def. 8.2.1) to a strictly upper triangular matrix (Ex. 8.2.4).

Notation 2.2.20. We denote by $N_n$ the $n \times n$ matrix defined by

$$N_n = \begin{pmatrix} 0 & 1 & & 0 \\ 0 & 1 & & 0 \\ & & \ddots & \ddots \\ 0 & & 0 & 1 \\ \end{pmatrix}.$$  \hfill (2.11)

That is, $N_n = (\alpha_{ij})$ where

$$\alpha_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}.$$  \hfill (2.12)

Exercise 2.2.21. Find $N_n^k$ for $k \geq 0$.

Fact 2.2.22. In Section 1.4 we defined the dot product of two vectors (Def. 1.4.1). The dot product may also be defined in terms of matrix multiplication. Let $x, y \in \mathbb{F}^k$. Then

$$x \cdot y = x^T y.$$  \hfill (2.13)

Exercise 2.2.23. If $v \perp 1$ then $Jv = 0$.

Notation 2.2.24. For a matrix $A \in \mathbb{F}^{k \times n}$, we sometimes write

$$A = [a_1 | \cdots | a_n]$$

where $a_i$ is the $i$-th column of $A$.

Exercise 2.2.25 (Extracting columns and elements of a matrix via multiplication). Let $e_i$ be the $i$-th column of $I$ (i.e., the $i$-th standard unit vector), and let $A = (\alpha_{ij}) = [a_1 | \cdots | a_n]$. Then

(a) $A e_j = a_j$;

(b) $e_i^T A e_j = \alpha_{ij}$.

Exercise 2.2.26 (Linear combination as a matrix product). Let $A = [a_1 | \cdots | a_n] \in \mathbb{F}^{k \times n}$ and let $x = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{F}^n$. Show that

$$Ax = \alpha_1 a_1 + \cdots + \alpha_n a_n.$$  \hfill (2.14)

Exercise 2.2.27 (Left multiplication acts column by column). Let $A \in \mathbb{F}^{k \times n}$ and let $B = [b_1 | \cdots | b_\ell] \in \mathbb{F}^{n \times \ell}$. (The $b_i$ are the columns of $B$.) Then

$$AB = [A b_1 | \cdots | A b_\ell].$$  \hfill (2.15)
2.2. MATRIX MULTIPLICATION

Proposition 2.2.28 (No cancellation). Let \( k \geq 2 \). Then for all \( n \) and for all \( x \in \mathbb{F}^n \), there exist \( k \times n \) matrices \( A \) and \( B \) (\( A \neq B \)) such that \( Ax = Bx \) but \( A \neq B \).

Proposition 2.2.29. Let \( A \in \mathbb{F}^{k \times n} \). If \( Ax = 0 \) for all \( x \in \mathbb{F}^n \), then \( A = 0 \).

Corollary 2.2.30 (Cancellation). If \( A, B \in \mathbb{F}^{k \times n} \) are matrices such that \( Ax = Bx \) for all \( x \in \mathbb{F}^n \), then \( A = B \). Note: compare with Prop. 2.2.28.

Proposition 2.2.31 (No double cancellation). Let \( k \geq 2 \). Then for all \( n \) and for all \( x \in \mathbb{F}^k \) and \( y \in \mathbb{F}^n \), there exist \( k \times n \) matrices \( A \) and \( B \) such that \( x^TAy = x^TBy \) but \( A \neq B \).

Definition 2.2.34 (Trace). The trace of a square matrix \( A = (\alpha_{ij}) \in M_n(\mathbb{F}) \) is the sum of its diagonal entries, that is,

\[
\text{Tr}(A) = \sum_{i=1}^{n} \alpha_{ii} \quad (2.15)
\]

Examples 2.2.35.

(a) \( \text{Tr}(I_n) = n \)

(b) \( \text{Tr}\begin{pmatrix} 3 & 1 & 2 \\ 4 & -2 & 4 \\ 1 & -3 & -2 \end{pmatrix} = -1 \)

Fact 2.2.36 (Linearity of the trace). (a) \( \text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B) \); (b) for \( \lambda \in \mathbb{F} \), \( \text{Tr}(\lambda A) = \lambda \text{Tr}(A) \); (c) for \( \alpha_i \in \mathbb{F} \), \( \text{Tr}(\sum \alpha_i A_i) = \sum \alpha_i \text{Tr} A_i \).

Exercise 2.2.37. Let \( v, w \in \mathbb{F}^n \). Then

\[
\text{Tr}(vw^T) = v^T w \quad (2.16)
\]

Note that \( vw^T \in M_n(\mathbb{F}) \) and \( v^T \in \mathbb{F} \).

Proposition 2.2.38. Let \( A \in \mathbb{F}^{k \times n} \) and let \( B \in \mathbb{F}^{n \times k} \). Then

\[
\text{Tr}(AB) = \text{Tr}(BA) \quad (2.17)
\]

Exercise 2.2.39. Show that the trace of a product is invariant under a cyclic permutation of the terms, i.e., if \( A_1, \ldots, A_k \) are matrices such that the product \( A_1 \cdots A_k \) is defined and is a square matrix, then

\[
\text{Tr}(A_1 \cdots A_k) = \text{Tr}(A_k A_1 \cdots A_{k-1}) \quad (2.18)
\]

Exercise 2.2.40. Show that the trace of a product is not invariant under all permutations of the terms. In particular, find \( 2 \times 2 \) matrices \( A, B, \) and \( C \) such that

\[
\text{Tr}(ABC) \neq \text{Tr}(BAC) \quad .
\]

Exercise 2.2.41 (Trace cancellation). Let \( B, C \in M_n(\mathbb{F}) \). Show that if \( \text{Tr}(AB) = \text{Tr}(AC) \) for all \( A \in M_n(\mathbb{F}) \), then \( B = C \).
2.3 Arithmetic of diagonal and triangular matrices

In this section we turn our attention to special properties of diagonal and triangular matrices.

Proposition 2.3.1. Let \( A = \text{diag}(\alpha_1, \ldots, \alpha_n) \) and \( B = \text{diag}(\beta_1, \ldots, \beta_n) \) be \( n \times n \) diagonal matrices and let \( \lambda \in \mathbb{F} \). Then

\[
A + B = \text{diag}(\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n) \tag{2.19}
\]
\[
\lambda A = \text{diag}(\lambda \alpha_1, \ldots, \lambda \alpha_n) \tag{2.20}
\]
\[
AB = \text{diag}(\alpha_1 \beta_1, \ldots, \alpha_n \beta_n) \tag{2.21}
\]

Proposition 2.3.2. Let \( A = \text{diag}(\alpha_1, \ldots, \alpha_n) \) be a diagonal matrix. Then \( A^k = \text{diag}(\alpha_1^k, \ldots, \alpha_n^k) \) for all \( k \).

Definition 2.3.3 (Substitution of a matrix into a polynomial). Let \( f \in \mathbb{F}[t] \) be the polynomial (\( \mathbb{F}[t] \) Def. 8.3.1) defined by

\[
f = \alpha_0 + \alpha_1 t + \cdots + \alpha_d t^d .
\]

Just as we may substitute \( \zeta \in \mathbb{F} \) for the variable \( t \) in \( f \) to obtain a value \( f(\zeta) \in \mathbb{F} \), we may also “plug in” the matrix \( A \in M_n(\mathbb{F}) \) to obtain \( f(A) \in M_n(\mathbb{F}) \). The only thing we have to be careful about is what we do with the scalar term \( \alpha_0 \); we replace it with \( \alpha_0 \) times the identity matrix, so

\[
f(A) := \alpha_0 I + \alpha_1 A + \cdots + \alpha_d A^d . \tag{2.22}
\]

Proposition 2.3.4. Let \( f \in \mathbb{F}[t] \) be a polynomial and let \( A = \text{diag}(\alpha_1, \ldots, \alpha_n) \) be a diagonal matrix. Then

\[
f(A) = \text{diag}(f(\alpha_1), \ldots, f(\alpha_n)) . \tag{2.23}
\]

In our discussion of the arithmetic of triangular matrices, we focus on the diagonal entries.

Notation 2.3.5. For the remainder of this section, the symbol * in a matrix will represent an arbitrary value with which we will not concern ourselves. We write

\[
\begin{pmatrix}
\alpha_1 & * \\
\alpha_2 & \ddots & \ddots \\
0 & \ddots & \ddots \\
\end{pmatrix}
\]

for

\[
\begin{pmatrix}
\alpha_1 & * & \cdots & * \\
\alpha_2 & \cdots & * \\
0 & \ddots & \ddots \\
\end{pmatrix}
\]

Proposition 2.3.6. Let

\[
A = \begin{pmatrix}
\alpha_1 & * \\
\alpha_2 & \ddots \\
0 & \ddots & \ddots \\
\end{pmatrix}
\]

and

\[
B = \begin{pmatrix}
\beta_1 & * \\
\beta_2 & \ddots \\
0 & \ddots & \ddots \\
\end{pmatrix}
\]
be upper triangular matrices and let $\lambda \in \mathbb{F}$. Then

$$A + B = \begin{pmatrix}
\alpha_1 + \beta_1 & \ast \\
\alpha_2 + \beta_2 & \\
0 & \ddots & \\
& & \alpha_n + \beta_n
\end{pmatrix}$$

(2.24)

$$\lambda A = \begin{pmatrix}
\lambda \alpha_1 & \ast \\
\lambda \alpha_2 & \\
0 & \ddots & \\
& & \lambda \alpha_n
\end{pmatrix}$$

(2.25)

$$AB = \begin{pmatrix}
\alpha_1 \beta_1 & \ast \\
\alpha_2 \beta_2 & \\
0 & \ddots & \\
& & \alpha_n \beta_n
\end{pmatrix}.$$  

(2.26)

**Proposition 2.3.7.** Let $A$ be as in Prop. 2.3.6. Then

$$A^k = \begin{pmatrix}
\alpha_1^k & \ast \\
\alpha_2^k & \\
0 & \ddots & \\
& & \alpha_n^k
\end{pmatrix}$$

(2.27)

for all $k$.

**Proposition 2.3.8.** Let $f \in \mathbb{F}[t]$ be a polynomial and let $A$ be as in Prop. 2.3.6. Then

$$f(A) = \begin{pmatrix}
f(\alpha_1) & \ast \\
f(\alpha_2) & \\
0 & \ddots & \\
& & f(\alpha_n)
\end{pmatrix}. $$

(2.28)

### 2.4 Permutation Matrices

**Definition 2.4.1 (Rook arrangement).** A rook arrangement is an arrangement of $n$ rooks on an $n \times n$ chessboard such that no pair of rooks attack each other. In other words, there is exactly one rook in each row and column.

**Fact 2.4.2.** The number of rook arrangements on an $n \times n$ chessboard is $n!$.

**Definition 2.4.3 (Permutation matrix).** A permutation matrix is a square matrix with the following properties.

(a) Every nonzero entry is equal to 1.

(b) Each row and column has exactly one nonzero entry.

Observe that rook arrangements correspond to permutation matrices where each rook is placed on a 1. Permutation matrices will be revisited in Chapter 6 where we discuss the determinant.

Recall that we denote by $[n]$ the set of integers $\{1, \ldots, n\}$.

**Definition 2.4.4 (Permutation).** A permutation is a bijection $\sigma : [n] \rightarrow [n]$. We write $\sigma : i \mapsto i^\sigma$.

**Example 2.4.5.** A permutation can be represented by a $2 \times n$ table like this.
This permutation takes 1 to 3, 2 to 6, 3 to 4, etc.

Note that any table obtained by rearranging columns represents the same permutation. For example, we may also represent the permutation \( \sigma \) by the table below:

\[
\begin{array}{cccccc}
    i & 4 & 6 & 1 & 5 & 2 \\
    i^\sigma & 1 & 2 & 3 & 5 & 6 \\
\end{array}
\]

Moreover, the permutation can be represented by a diagram, where the arrow \( i \to i^\sigma \) means that \( i \) maps to \( i^\sigma \).

\[\text{FIGURE HERE}\]

**Definition 2.4.6 (Composition of permutations).** Let \( \sigma, \tau : [n] \to [n] \) be permutations. The composition of \( \tau \) with \( \sigma \), denoted \( \sigma \tau \), is the permutation which maps \( i \) to \( i^{\sigma \tau} \) defined by

\[
i^{\sigma \tau} := (i^\sigma)\tau . \tag{2.29}\]

This may be represented as

\[
i \xrightarrow{\sigma} i^\sigma \xrightarrow{\tau} i^{\sigma \tau} .
\]

**Example 2.4.7.** Let \( \sigma \) be the permutation of Example 2.4.5 and let \( \tau \) be the permutation given in the table

\[
\begin{array}{cccccc}
    i & 1 & 2 & 3 & 4 & 5 \\
    i^\tau & 3 & 6 & 4 & 1 & 5 \\
\end{array}
\]

Then we can find the table representing the permutation \( \sigma \tau \) by rearranging the table for \( \tau \) so that its first row is in the same order as the second row of the table for \( \sigma \), i.e.,

\[
\begin{array}{cccccc}
    i & 3 & 6 & 4 & 1 & 5 \\
    i^\tau & 6 & 3 & 5 & 4 & 2 \\
\end{array}
\]

and then combining the two tables:

\[
\begin{array}{cccccc}
    i & 1 & 2 & 3 & 4 & 5 & 6 \\
    i^\sigma & 3 & 6 & 4 & 1 & 5 & 2 \\
    i^{\sigma \tau} & 6 & 3 & 5 & 4 & 2 & 1 \\
\end{array}
\]

That is, the table corresponding to the permutation \( \sigma \tau \) is

\[
\begin{array}{cccccc}
    i & 1 & 2 & 3 & 4 & 5 & 6 \\
    i^{\sigma \tau} & 6 & 3 & 5 & 4 & 2 & 1 \\
\end{array}
\]

**Definition 2.4.8 (Identity permutation).** The identity permutation \( \text{id} : [n] \to [n] \) is the permutation defined by \( i^{\text{id}} = i \) for all \( i \in [n] \).

**Definition 2.4.9 (Inverse of a permutation).** Let \( \sigma : [n] \to [n] \) be a permutation. Then the inverse of \( \sigma \) is the permutation \( \sigma^{-1} \) such that \( \sigma \sigma^{-1} = \text{id} \). So \( \sigma^{-1} \) takes \( i^{\sigma} \) to \( i \).

**Example 2.4.10.** To find the table corresponding to the inverse of a permutation \( \sigma \), we first switch the rows of the table corresponding to \( \sigma \) and then rearrange the columns so that they are in natural order. For example, when we switch the rows of the table in Example 2.4.5, we have

\[
\begin{array}{cccccc}
    i & 3 & 6 & 4 & 1 & 5 \\
    i^{-1} & 1 & 2 & 3 & 4 & 5 \\
\end{array}
\]

That is, the table corresponding to the permutation \( \sigma^{-1} \) is

\[
\begin{array}{cccccc}
    i & 3 & 6 & 4 & 1 & 5 & 2 \\
    i^{-1} & 1 & 2 & 3 & 4 & 5 & 6 \\
\end{array}
\]
Rearranging the columns so that they are in natural order gives us the table

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>i^σ</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**Definition 2.4.11.** Let \( \sigma : [n] \to [n] \) be a permutation. The permutation matrix corresponding to \( \sigma \), denoted \( P_\sigma \), is the \( n \times n \) matrix whose \( (i,j) \) entry is 1 if \( \sigma(i) = j \) and 0 otherwise.

**Exercise 2.4.12.** Let \( \sigma, \tau : [n] \to [n] \) be permutations, let \( A \in \mathbb{F}^{k \times n} \), and let \( B \in \mathbb{F}^{n \times \ell} \).

(a) What is \( AP_\sigma \)?
(b) What is \( P_\sigma B \)?
(c) What is \( P_{\sigma^{-1}} \)?
(d) Show that \( P_\sigma P_\tau = P_{\tau \sigma} \) (note the conflict in conventions for composition of permutations and matrix multiplication).

**2.5 Additional exercises**

**Exercise 2.5.1.** Let \( A \) be a matrix. Show, without calculation, that \( A^T A \) is symmetric.

**Definition 2.5.2 (Commutator).** Let \( A \in \mathbb{F}^{k \times n} \) and \( A \in \mathbb{F}^{n \times k} \). Then the commutator of \( A \) and \( B \) is \( AB - BA \).

**Definition 2.5.3.** Two matrices \( A, B \in M_n(\mathbb{F}) \) commute if \( AB = BA \).

**Exercise 2.5.4.** (a) Find an example of two \( 2 \times 2 \) matrices that do not commute.

(b) (Project) Interpret and prove the following statement:

Almost all pairs of \( 2 \times 2 \) integral matrices (matrices in \( M_2(\mathbb{Z}) \)) do not commute.

**Exercise 2.5.5.** Let \( D \in M_n(\mathbb{F}) \) be a diagonal matrix such that all diagonal entries are distinct. Show that if \( A \in M_n(\mathbb{F}) \) commutes with \( D \) then \( A \) is a diagonal matrix.

**Exercise 2.5.6.** Show that only the scalar matrices (Def. 2.2.13) commute with all matrices in \( M_n(\mathbb{F}) \). (A scalar matrix is a matrix of the form \( \lambda I \).)

**Exercise 2.5.7.**

(a) Show that the commutator of two matrices over \( \mathbb{R} \) is never the identity matrix.

(b) Find \( A, B \in M_p(\mathbb{F}_p) \) such that their commutator is the identity.

**Exercise 2.5.8 (Submatrix sum).** Let \( I_1 \subseteq [k] \) and \( I_2 \subseteq [n] \), and let \( B \) be the submatrix (Def. 3.3.11) of \( A \in \mathbb{F}^{k \times n} \) with entries \( a_{ij} \) for \( i \in I_1, j \in I_2 \). Find vectors \( a \) and \( b \) such that \( a^T Ab \) equals the sum of the entries of \( B \).

**Definition 2.5.9 (Vandermonde matrix).** The **Vandermonde matrix** generated by \( \alpha_1, \ldots, \alpha_n \)
is the $n \times n$ matrix
\[
V(\alpha_1, \ldots, \alpha_n) = \begin{pmatrix}
1 & 1 & \cdots & 1 \\
\alpha_1 & \alpha_2 & \cdots & \alpha_n \\
\alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1}
\end{pmatrix}.
\] (2.30)

Exercise 2.5.10. Let $A$ be a Vandermonde matrix generated by distinct $\alpha_i$. Show that the rows of $A$ are linearly independent. Do not use determinants.

Exercise 2.5.11. Prove that polynomials of a matrix commute: let $A$ be a square matrix and let $f, g \in \mathbb{F}[t]$. Then $f(A)$ and $g(A)$ commute. In particular, $A$ commutes with $f(A)$.

Definition 2.5.12 (Circulant matrix). The circulant matrix generated by the sequence $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ of scalars is the $n \times n$ matrix
\[
C(\alpha_0, \alpha_1, \ldots, \alpha_{n-1}) = \begin{pmatrix}
\alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} \\
\alpha_{n-1} & \alpha_0 & \cdots & \alpha_{n-2} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1 & \alpha_2 & \cdots & \alpha_0
\end{pmatrix}.
\] (2.31)

Exercise 2.5.13. Prove that all circulant matrices commute. Prove this
(a) directly,
(b) in a more elegant way, by showing that all circulant matrices are polynomials of a particular circulant matrix.

Definition 2.5.14 (Jordan block). For $\lambda \in \mathbb{C}$ and $n \geq 1$, the Jordan block $J(n, \lambda)$ is the matrix
\[
J(n, \lambda) := \lambda I + N_n
\] (2.32)
where $N_n$ is the matrix defined in Notation 2.2.20.

Exercise 2.5.15. Let $f \in \mathbb{C}[t]$. Prove that $f(J(n, \lambda))$ is the matrix
\[
\begin{pmatrix}
f(\lambda) & f'(\lambda) & \frac{1}{2!}f^{(2)}(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\
f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\
\vdots & \vdots & \ddots & \vdots \\
0 & f(\lambda) & f'(\lambda) & f(\lambda)
\end{pmatrix}.
\] (2.33)

Exercise 2.5.16. The converse of the second statement in Ex. 2.5.11 would be:

(*) The only matrices that commute with $A$ are the polynomials of $A$.

(a) Find a matrix $A$ for which (*) is false.
(b) For which diagonal matrices is (*) true?
(c) Prove: (*) is true for Jordan blocks.
(d) Characterize the matrices over $\mathbb{C}$ for which (*) is true, in terms of their Jordan blocks (1.10 2.?)
2.5. ADDITIONAL EXERCISES

Project 2.5.17. For $A \in M_n(\mathbb{R})$, let $f(A, k)$ denote the largest absolute value of all entries of $A^k$, and define

$$M_n^{(\lambda)}(\mathbb{R}) := \{A \in M_n(\mathbb{R}) \mid f(A, 1) \leq \lambda\}$$

(2.34)

(the matrices where all entries have absolute value $\leq \lambda$). Define

$$f_1(n, k) = \max_{A \in M_n^{(1)}(\mathbb{R})} f(A, k),$$

$$f_2(n, k) = \max_{A \in M_n^{(1)}(\mathbb{R}) \text{ nilpotent}} f(A, k),$$

$$f_3(n, k) = \max_{A \in M_n^{(1)}(\mathbb{R}) \text{ strictly upper triangular}} f(A, k).$$

Find the rate of growth of these functions in terms of $k$ and $n$. 
Chapter 3
(F) Matrix Rank

3.1 Column and row rank

In Section 1.3, we defined the rank of a list of column or row vectors. We now view this concept in terms of the columns and rows of a matrix.

Definition 3.1.1 (Column- and row-rank). The column-rank of a matrix is the rank of the list of its column vectors. The row-rank of a matrix is the rank of the list of its row vectors. A matrix $A \in \mathbb{F}^{k \times n}$ is of full row-rank if its row-rank is $k$. $A$ is of full column-rank if its column-rank is $n$.

Exercise 3.1.5. Let $A$ and $B$ be $k \times n$ matrices. Then
\[|\text{rk}_\text{col} A - \text{rk}_\text{col} B| \leq \text{rk}_\text{col}(A+B) \leq \text{rk}_\text{col} A + \text{rk}_\text{col} B.\]

Exercise 3.1.6. Let $A$ be a $k \times n$ matrix and let $B$ be an $n \times \ell$ matrix. Then \[\text{col-space}(AB) \leq \text{col-space} A.\]

Exercise 3.1.7. Let $A$ be a $k \times n$ matrix and let $B$ be an $n \times \ell$ matrix. Then \[\text{rk}_\text{col}(AB) \leq \text{rk}_\text{col} A.\]

3.2 Elementary operations and Gaussian elimination

There are three kinds of “elementary operations” on the columns of a matrix and three corresponding elementary operations on the rows of a matrix. The significance of these operations is that while they can simplify the matrix by turning most of its elements into zero, they do not change the rank of the matrix, and col-space $A \leq \mathbb{F}^k$ and row-space $A \leq \mathbb{F}^n$.

Proposition 3.1.3. The column-rank of a matrix is equal to the dimension of its column space, and the row-rank of a matrix is equal to the dimension of its row space.

Definition 3.1.4 (Full column- and row-rank). A $k \times n$ matrix $A$ is of full column-rank if its column-rank is $k$. $A$ is of full row-rank if its row-rank is $n$. 
they change the value of the determinant in an easy-to-follow manner.

We call the three column operations column shearing, column scaling, and column swapping, and use the corresponding terms for rows as well. The most powerful of these is “column shearing.” While the literature sometimes calls this operation “column addition,” I find this term misleading: we don’t “add columns” – that is a term more adequately applied in the context of the multilinearity of the determinant. I propose the term “column shearing” from physical analogy which I will explain below. Similarly misleading is the term “column multiplication” often used to describe what I propose to call “column scaling.” The term “column multiplication” seems to suggest that we multiply columns, as we do in the definition of an orthogonal matrix. Our “column swapping” is often described as “column switching.” Although “switching” seems to be an adequate term, I believe “swapping” better describes what is actually happening.

Let \( A = [a_1 | \cdots | a_n] \) be a matrix; the \( v_i \) are the columns of this matrix. The definitions below refer to this matrix.

**Definition 3.2.1 (Column shearing).** The column shearing operation, denoted by \( \text{shear}_c(i,j,\lambda) \), takes three parameters: column indices \( i \neq j \) and a scalar \( \lambda \). The operation replaces \( a_i \) by \( a_i + \lambda a_j \).

Note that the only column that changes as a result of the \( \text{shear}_c(i,j,\lambda) \) operation is column \( i \). Remember the condition \( i \neq j \).

**Definition 3.2.2 (Column scaling).** The column scaling operation, denoted by \( \text{scale}_c(i,\lambda) \), takes two parameters: a column index \( i \) and a scalar \( \lambda \neq 0 \). The operation replaces \( a_i \) by \( \lambda a_i \).

Note that the only column that changes as a result of the \( \text{scale}_c(i,\lambda) \) operation is column \( i \). Remember the condition \( \lambda \neq 0 \).

**Definition 3.2.3 (Column swapping).** The column swapping operation, denoted by \( \text{swap}_c(i,j) \), takes two parameters: column indices \( i \neq j \). The operation swaps the columns \( a_i \) and \( a_j \).

So what happens can be described with reference to a temporary vector \( z \) as the following sequence of operations:

\[
z \leftarrow v_i, \quad v_i \leftarrow v_j, \quad v_j \leftarrow z.
\]

Note that the only columns that change as a result of the \( \text{swap}_c(i,j) \) operation are columns \( i \) and \( j \). Remember the condition \( i \neq j \).

**Definition 3.2.4. Elementary row operations**

The elementary row operations \( \text{shear}_r(i,j,\lambda) \), \( \text{scale}_r(i,\lambda) \), and \( \text{swap}_r(i,j) \) are defined analogously.

**Exercise 3.2.5.** Each of the elementary operations is invertible and the inverse of each elementary operation is an elementary operation of the same type. Specifically, the inverse of \( \text{shear}_c(i,j,\lambda) \) is \( \text{shear}_c(i,j,-\lambda) \); the inverse of \( \text{scale}_c(i,\lambda) \) is of \( \text{scale}_c(i,\lambda^{-1}) \), and \( \text{swap}_c(i,j) \) is its own inverse. The analogous statements hold for the elementary row operations.
Definition 3.2.6 (Elementary matrix). An elementary matrix is a square matrix that differs from the identity matrix in an elementary operation.

Accordingly, we shall have three kinds of elementary matrices: shearing, scaling, and swapping matrices.

Definition 3.2.7 (Shearing matrix). Let $i \neq j$, $1 \leq i, j \leq n$. We denote by $E_{ij}$ the $n \times n$ matrix which has 1 in the $(i, j)$ position and 0 in every other position. A shearing matrix $B$ is an $n \times n$ matrix of the form $B = I + \lambda E_{ij}$ for $\lambda \in \mathbb{F}$.

Exercise 3.2.8. Let $A \in \mathbb{F}^{k \times n}$ be a $k \times n$ matrix, $\lambda \in \mathbb{F}$, and $B$ the $n \times n$ shearing matrix $B = I + \lambda E_{ij}$. Show that $AB$ is the matrix obtained from $A$ by performing the operation shear$_c(j, i, \lambda)$. Infer (do not repeat the same argument) that $BA$ is obtained from $A$ by performing the row operation shear$_r(i, j, \lambda)$.

Remark 3.2.9. In physics, “shearing forces” are “unaligned forces pushing one part of a body in one specific direction, and another part of the body in the opposite direction.” (Wikipedia) A pair of scissors exerts shearing forces on the material it tries to cut. A nice example is a deck of cards on a horizontal table being pushed horizontally on the top while being held in place at the bottom, causing the cards to slide. Thus the deck, which initially has a rectangle as its vertical cross-section, ends up with a parallelogram of the same height as its vertical cross-section. The distance of each card from the base remains constant, and the distance each card slides is proportional to its distance from the base. This is precisely what an application of a shearing matrix to the deck achieves in 2D or 3D.

Definition 3.2.10 (Scaling matrix). Let $1 \leq i \leq n$ and $\lambda \in \mathbb{F}$, $\lambda \neq 0$. We denote by $D(i, \lambda)$ the $n \times n$ diagonal matrix which has $\lambda$ in the $(i, i)$ position and 1 in all the other diagonal positions.

Exercise 3.2.11. Let $A \in \mathbb{F}^{k \times n}$ be a $k \times n$ matrix, $\lambda \in \mathbb{F}$, $\lambda \neq 0$, and $1 \leq i \leq n$. Show that $AD(i, \lambda)$ is the matrix obtained from $A$ by performing the column operation scale$_c(i, \lambda)$. Infer (do not repeat the same argument) that $D(i, \lambda)A$ is obtained from $A$ by performing the row operation scale$_r(i, \lambda)$.

Definition 3.2.12 (Swapping matrix). Let $1 \leq i, j \leq n$ where $i \neq j$. We denote by $T_{ij}$ the $n \times n$ matrix obtained from the identity matrix by swapping the column $i$ and $j$. Note that by swapping rows $i$ and $j$ we obtain the same matrix.

Exercise 3.2.13. Let $A \in \mathbb{F}^{k \times n}$ be a $k \times n$ matrix, $1 \leq i, j \leq n$, $i \neq j$. Show that $A \cdot T_{ij}$ is the matrix obtained from $A$ by performing the column operation swap$_c(i, j)$. Infer (do not repeat the same argument) that $T_{ij}A$ is obtained from $A$ by performing the row operation swap$_r(i, j)$.

Exercise 3.2.14. Let $A$ be a matrix with linearly independent rows. Then by performing a
3.3. THE SECOND MIRACLE OF LINEAR ALGEBRA

A series of row shearing, row swapping, and column swapping operations we can bring the matrix to the form $[D \mid B]$ where $D$ is a diagonal matrix no zeros in the diagonal.

**Exercise 3.2.15.** Let $A$ be a $k \times n$ matrix. Then by performing a series of row shearing, row swapping, and column swapping operations we can bring the $A$ to the $2 \times 2$ block-matrix form

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

where the top left block, $D$, is an $r \times r$ diagonal matrix with no zeros in the diagonal (for some value $r$ which will turn out to be the rank of $A$) and the $k - r$ bottom rows are zero.

The process of transforming a matrix into a matrix of the form described in Ex. 3.2.15 by performing a series of row shearing, row swapping, column shearing, and column swapping operations is called **Gauss-Jordan elimination**.

**Exercise 3.2.16.** Any matrix can be transformed into an upper triangular matrix by performing a series of row shearing, row swapping, and column swapping operations.

The process of transforming a matrix into a matrix of the form described in Ex. 3.2.15 by performing a series of row shearing, row swapping, and column swapping operations is called **Gaussian elimination**.

**Exercise 3.2.17.** Let $A$ be a $k \times n$ matrix. Then by performing a series of row shearing, row swapping, column shearing, and column swapping operations we can bring the $A$ to the $2 \times 2$ block-matrix form

$$
\begin{pmatrix}
D & 0 \\
0 & 0
\end{pmatrix}
$$

where the top left block, $D$, is an $r \times r$ diagonal matrix with no zeros in the diagonal (for some value $r$ which will turn out to be the rank of $A$); the rest of the matrix is zero.

The main result of this section is the following theorem.

**Theorem 3.3.1** (Second Miracle of Linear Algebra). The row-rank of a matrix is equal to its column-rank.  

In order to prove this theorem, we will first need to prove the following two lemmas.

**Lemma 3.3.2.** Elementary column operations do not change the column-rank of a matrix.

The proof of the next lemma is somewhat more involved; we will break its proof into exercises.

**Lemma 3.3.3.** Elementary row operations do not change the column-rank of a matrix.
Exercise 3.3.4. Use these two lemmas together with Ex. 3.2.17 to prove Theorem 3.3.1.

In order to complete our proof of the Second Miracle, we now only need to prove Lemma 3.3.3.

The following exercise demonstrates why the proof of Lemma 3.3.3 is not as straightforward as the proof of Lemma 3.3.2.

Exercise 3.3.5. An elementary row operation can change the column space of a matrix.

Proposition 3.3.6. Let \( A = [a_1 | \cdots | a_n] \in \mathbb{F}^{k \times n} \) be a matrix with columns \( a_1, \ldots, a_n \). Let
\[
\sum_{i=1}^{n} \alpha_i a_i = 0
\]
be a linear relation among the columns. Let \( A' = [a'_1 | \cdots | a'_n] \) be the result of applying an elementary row operation to \( A \). Then the columns of \( A' \) obey the same linear relation, that is,
\[
\sum_{i=1}^{n} \alpha_i a'_i = 0.
\]

Corollary 3.3.7. If the columns \( v_{i_1}, \ldots, v_{i_k} \) are linearly independent, then this remains true after an elementary row operation.

Exercise 3.3.8. Complete the proof of Lemma 3.3.3.

Because the Second Miracle of Linear Algebra establishes that the row-rank and column-rank of a matrix \( A \) are equal, it is no longer necessary to differentiate between them; this quantity is simply referred to as the rank of \( A \), denoted \( \text{rk} A \).

Definition 3.3.9 (Full rank). Let \( A \in M_n(\mathbb{F}) \) be a matrix. Then \( A \) is of full rank if \( \text{rk} A = n \).

Observe that full rank is the same as full column- or row-rank (Def. 3.1.4) only in the case of square matrices.

Definition 3.3.10 (Nonsingular matrix). An \( n \times n \) matrix is nonsingular if it is of full rank, i.e., \( \text{rk} A = n \).

Definition 3.3.11 (Submatrix). Let \( A \in \mathbb{F}^{k \times n} \) be a matrix. Then the matrix \( B \) is a submatrix of \( A \) if it can be obtained by deleting rows and columns from \( A \).

In other words, a submatrix of a matrix \( A \) is a matrix obtained by taking the intersection of a set of the rows of \( A \) with a set of the columns of \( A \).

Theorem 3.3.12 (Rank vs. nonsingular submatrices). Let \( A \in \mathbb{F}^{k \times n} \) be a matrix. Then \( \text{rk} A \) is the largest value of \( r \) such that \( A \) has a nonsingular \( r \times r \) submatrix. \( \diamond \)

Exercise 3.3.13. Show that for all \( k \), the intersection of \( k \) linearly independent rows with \( k \) linearly independent columns can be singular. In fact, it can be the zero matrix.
Exercise 3.3.14. Let \( A \) be a matrix of rank \( r \). Show that the intersection of any \( r \) linearly independent rows with any \( r \) linearly independent columns is a nonsingular \( r \times r \) submatrix of \( A \). (Note: this exercise is more difficult than Theorem 3.3.12 and is not needed for the proof of Theorem 3.3.12).

Exercise 3.3.15. Let \( A \) be a matrix. Show that if the intersection of \( k \) linearly independent columns with \( \ell \) linearly independent rows of \( A \) has rank \( s \), then \( \text{rk}(A) \geq k + \ell - s \).

3.4 Matrix rank and invertibility

Definition 3.4.1 (Left and right inverse). Let \( A \in \mathbb{F}^{k \times n} \). The matrix \( B \in \mathbb{F}^{n \times k} \) is a left inverse of \( A \) if \( BA = I_n \). Likewise, the matrix \( C \in \mathbb{F}^{n \times k} \) is a right inverse of \( A \) if \( AC = I_k \).

Proposition 3.4.2. Let \( A \in \mathbb{F}^{k \times n} \).

(a) Show that \( A \) has a right inverse if and only if \( A \) has full row-rank, i.e., \( \text{rk} A = k \).

(b) Show that \( A \) has a left inverse if and only if \( A \) has full column-rank, i.e., \( \text{rk} A = n \).

Note in particular that if \( A \) has a right inverse, then \( k \leq n \), and if \( A \) has a left inverse, then \( k \geq n \).

Corollary 3.4.3. Let \( A \) be a nonsingular square matrix. Then \( A \) has both a right and a left inverse.

Exercise 3.4.4. For all \( k < n \), find a \( k \times n \) matrix that has infinitely many right inverses.

Definition 3.4.5 (Two-sided inverse). Let \( A \in \mathbb{F}^{k \times n} \). Then the matrix \( B \in \mathbb{F}^{n \times k} \) is a (two-sided) inverse of \( A \) if \( AB = BA = I_n \). The inverse of \( A \) is denoted \( A^{-1} \). If \( A \) has an inverse, then \( A \) is said to be invertible.

Proposition 3.4.6. Let \( A \) be a matrix. If \( A \) has a left inverse as well as a right inverse, then \( A \) has a unique two-sided inverse and it has no left or right inverse other than the two-sided inverse.

The proof of this lengthy statement is just one line, based solely on the associativity of matrix multiplication. The essence of the proof is in the next lemma.

Lemma 3.4.7. Let \( A \in \mathbb{F}^{k \times n} \) be a matrix with a right inverse \( B \) and a left inverse \( C \). Then \( B = C \) is a two-sided inverse of \( A \) and \( k = n \). \( \diamond \)

Corollary 3.4.8. Under the conditions of Lemma 3.4.7, \( k = n \) and \( B = C \) is a two-sided inverse. Moreover, if \( C_1 \) is also a left inverse, then \( C_1 = C \); analogously, if \( B_1 \) is also a right inverse, then \( B_1 = B \).

Corollary 3.4.9. Let \( A \) be a matrix with a left inverse. Then \( A \) has at most one right inverse.

Corollary 3.4.10. A matrix \( A \) has an inverse if and only if \( A \) is a nonsingular square matrix.

Corollary 3.4.11. If \( A \) has both a right and a left inverse then \( k = n \).
Theorem 3.4.12. For $A \in M_n(F)$, the following are equivalent.

(a) $\text{rk} A = n$

(b) $A$ has a right inverse

(c) $A$ has a left inverse

(d) $A$ has an inverse

Exercise 3.4.13. Assume $F$ is infinite and let $A \in F^{k \times n}$ where $n > k$. If $A$ has a right inverse, then $A$ has infinitely many right inverses.

3.5 Codimension (optional)

The reader comfortable with abstract vector spaces can skip to Chapter ?? for a more general discussion of this material.

Definition 3.5.1 (Codimension). Let $W \leq F^n$. The codimension of $W$ is the minimum number of vectors that together with $W$ generate $F^n$ and is denoted by $\text{codim} W$ or $\text{codim} W$. (Note that this is the dimension of the quotient space $F^n/W$; see Def. ??)

Proposition 3.5.2. If $W \leq F^n$ then $\text{codim} W = n - \dim W$.

Proposition 3.5.3 (Dual modular equation). Let $U, W \leq F^n$. Then

$$\text{codim}(U + W) + \text{codim}(U \cap W) = \text{codim} U + \text{codim} W. \quad (3.2)$$

Corollary 3.5.4. Let $U, W \leq F^n$. Then

$$\text{codim}(U \cap W) \leq \text{codim} U + \text{codim} W. \quad (3.3)$$

Corollary 3.5.5. Let $U, W \leq F^n$ with $\dim U = r$ and $\text{codim} W = t$. Then

$$\dim(U \cap W) \geq \max\{r - t, 0\}. \quad (3.4)$$

In Section [1.3] we defined the rank of a set of column vectors (Def. [1.3.28]) and then showed this to be equal to be the dimension of their span (Cor. [1.3.46]). We now define the corank of a set of vectors.

Definition 3.5.6 (Corank). The corank of a set $S \subseteq F^n$, denoted $\text{corank} S$, is defined by

$$\text{corank} S := \text{codim} (\text{span} S). \quad (3.5)$$

Definition 3.5.7 (Null space). The null space or kernel of a matrix $A \in F^{k \times n}$, denoted $\text{null}(A)$, is the set

$$\text{null}(A) = \{v \in F^n \mid Av = 0\}. \quad (3.6)$$

Exercise 3.5.8. Let $A \in F^{k \times n}$. Show that

$$\text{rk} A = \text{codim}(\text{null}(A)). \quad (3.7)$$

Proposition 3.5.9. Let $U \leq F^n$ and let $W \leq F^k$ such that $\dim W = \text{codim} U = \ell$. Then there is a matrix $A \in F^{\ell \times k}$ such that $\text{null}(A) = U$ and col-space $A = W$. 
3.6. ADDITIONAL EXERCISES

**Definition 3.5.10** (Corank of a matrix). Let $A \in \mathbb{F}^{k \times n}$. We define the corank of $A$ as the corank of its column space, i.e.,

$$\text{corank } A := k - \text{rk } A.$$ (3.8)

**Exercise 3.5.11.** When is $\text{corank } A = \text{corank } A^T$?

**Exercise 3.5.12.** Let $A \in \mathbb{F}^{k \times n}$ and let $B \in \mathbb{F}^{n \times \ell}$. Show that

$$\text{corank}(AB) \leq \text{corank } A + \text{corank } B.$$ (3.9)

### 3.6 Additional exercises

**Exercise 3.6.1.** Let $A = [v_1 \mid \cdots \mid v_n]$ be a matrix. True or false: if the columns $v_1, \ldots, v_\ell$ are linearly independent this remains true after performing elementary column operations on $A$.

**Proposition 3.6.2.** Let $A$ be a matrix with at most one nonzero entry in each row and column. Then $\text{rk } A$ is the number of nonzero entries in $A$.

**Numerical exercise 3.6.3.** Perform a series of elementary row operations to determine the rank of the matrix

$$
\begin{pmatrix}
1 & 3 & 2 \\
-5 & -2 & 3 \\
-3 & 4 & 7
\end{pmatrix}.
$$

*Self-check:* use a different sequence of elementary operations and verify that you obtain the same answer.

**Exercise 3.6.4.** Determine the ranks of the $n \times n$ matrices

(a) $A = (\alpha_{ij})$ where $\alpha_{ij} = i + j$

(b) $B = (\beta_{ij})$ where $\beta_{ij} = ij$

(c) $C = (\gamma_{ij})$ where $\gamma_{ij} = i^2 + j^2$

**Proposition 3.6.5.** Let $A \in \mathbb{F}^{k \times \ell}$ and $B \in \mathbb{F}^{\ell \times m}$. Then

$$\text{rk}(AB) \leq \min\{\text{rk}(A), \text{rk}(B)\}.$$ 

**Proposition 3.6.6.** Let $A, B \in \mathbb{F}^{k \times \ell}$. Then

$$\text{rk}(A + B) \leq \text{rk}(A) + \text{rk}(B).$$

**Exercise 3.6.7.** Find an $n \times n$ matrix $A$ of rank $n - 1$ such that $A^n = 0$.

**Exercise 3.6.8.** Let $A \in \mathbb{F}^{k \times n}$. Then $\text{rk } A$ is the smallest $r$ such that there exist matrices $B \in \mathbb{F}^{k \times r}$ and $C \in \mathbb{F}^{r \times n}$ with $A = BC$.

**Exercise 3.6.9.** Show that the matrix $A$ is of rank $r$ if and only if $A$ can be expressed as the sum of $r$ matrices of rank 1.

**Exercise 3.6.10** (Characterization of matrices of rank 1). Let $A \in \mathbb{F}^{k \times n}$. Show that $\text{rk } A = 1$ if and only if there exist column vectors $a \in \mathbb{F}^k$ and $b \in \mathbb{F}^n$ such that $A = ab^T$. 
Let \( A \in \mathbb{F}^{k \times n} \) and let \( \mathbb{F} \) be a subfield of the field \( \mathbb{G} \). Then we can also view \( A \) as a matrix over \( \mathbb{G} \). However, this changes the notion of linear combinations and therefore in principle it could affect the rank. We shall see that this is not the case, but in order to be able to reason about this question, we temporarily use the notation \( \text{rk}_\mathbb{F}(A) \) and \( \text{rk}_\mathbb{G}(A) \) to denote the rank of \( A \) with respect to the corresponding fields. We will also write \( \text{rk}_\mathbb{p} \) to mean \( \text{rk}_{\mathbb{F}_p}(A) \).

**Lemma 3.6.11.** Being nonsingular is insensitive to field extensions. \( \diamondsuit \)

**Corollary 3.6.12.** Let \( \mathbb{F} \) be a subfield of \( \mathbb{G} \), and let \( A \in \mathbb{F}^{k \times n} \). Then

\[
\text{rk}_\mathbb{F}(A) = \text{rk}_\mathbb{G}(A).
\]

**Exercise 3.6.13.** Let \( A \in \mathbb{R}^{k \times n} \) be a matrix with integer coefficients. Then the entries of \( A \) can also be considered residue classes of \( \mathbb{Z} \), that is, we can instead consider \( A \) as an element of \( \mathbb{F}_p^{k \times n} \). Observe that \( \mathbb{F}_p \) is not a subfield of \( \mathbb{R} \), because the operations of addition and multiplication are not the same.

(a) Show that \( \text{rk}_\mathbb{p}(A) \leq \text{rk}(A) \).

(b) For every \( p \), find a \((0,1)\) matrix \( A \) (that is, a matrix whose entries are only 0 and 1) where this inequality is strict, i.e., \( \text{rk}_\mathbb{p}(A) < \text{rk}(A) \).

**Proposition 3.6.14.**

(a) Let \( A \) be a matrix with real entries. Then \( \text{rk}(A^T A) = \text{rk}(A) \).

(b) This is false over \( \mathbb{C} \) and over \( \mathbb{F}_p \).

**Exercise 3.6.15.**

(a) Show that \( \text{rk}(J_n - I_n) = n \).

(b) Show that

\[
\text{rk}_2(J_n - I_n) = \begin{cases} 
  n & \text{n even} \\
  n - 1 & \text{n odd} \end{cases}
\]

**Exercise 3.6.16.** Let \( A = (\alpha_{ij}) \) be a matrix of rank \( r \), and let \( A' \) be the matrix obtained by squaring every element of \( A \).

(a) Let \( A' = (\alpha_{ij}^2) \) be the matrix obtained by squaring every element of \( A \). Show that \( \text{rk}(A') \leq \binom{r+1}{2} \).

(b) Now let \( A' = (\alpha_{ij}^d) \). Show that \( \text{rk}(A') \leq \binom{r+d-1}{d} \).

(c) Let \( f \) be a polynomial of degree \( d \), and let \( A_f = (f(\alpha_{ij})) \). Prove that \( \text{rk}(A_f) \leq \binom{r+d}{d} \).

(d) Show that each of these bounds are tight for all \( r \), i.e., for every \( r \) and \( d \) there exists a matrix \( A \) such that \( \text{rk}(A') = \binom{r+d-1}{d} \), and for every \( r \) and \( f \) there exists a matrix \( B \) such that \( \text{rk}(B_f) = \binom{r+d}{d} \).
Chapter 4

(\mathbb{F}) Theory of Systems of Linear Equations I: Qualitative Theory

4.1 Homogeneous systems of linear equations

Matrices allow us to concisely express systems of linear equations. In particular, consider the general system of \( k \) linear equations in \( n \) unknowns,

\[
\begin{align*}
\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n &= \beta_1 \\
\alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n &= \beta_2 \\
&\vdots \\
\alpha_{k1}x_1 + \alpha_{k2}x_2 + \cdots + \alpha_{kn}x_n &= \beta_k
\end{align*}
\]

Here, the \( \alpha_{ij} \) and \( \beta_i \) are scalars, while the \( x_j \) are unknowns. In Section 1.1 we represented this system as a linear combination of column vectors. Matrices allow us to write this system even more concisely as \( Ax = b \), where

\[
A = \begin{pmatrix}
\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\
\alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{k1} & \alpha_{k2} & \cdots & \alpha_{kn}
\end{pmatrix}, \quad x = \begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{pmatrix}, \quad b = \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_k
\end{pmatrix}. \quad (4.1)
\]

We know that the simplest linear equation is of the form \( ax = b \) (one equation in one unknown); remarkably, the notation does not become any more complex when we have a general system of linear equations. The first question we ask about any system of equations is its solvability.

Definition 4.1.1 (Solvable system of linear equations). Given a matrix \( A \in \mathbb{F}^{k \times n} \) and a vector \( b \in \mathbb{F}^k \), we say that the system \( Ax = b \) of linear equations is solvable if there exists a vector \( x \in \mathbb{F}^n \) that satisfies \( Ax = b \).

Definition 4.1.2 (Homogeneous system of linear equations). The system \( Ax = 0 \) is called a homogeneous system of linear equations. Every system of homogeneous linear equations is solvable.

Definition 4.1.3 (Trivial solution to a homogeneous system of linear equations). The trivial solution to the homogeneous system of linear equations \( Ax = 0 \) is the solution \( x = 0 \).
So when presented with a homogeneous system of linear equations, the question we ask is not, “Is this system solvable?” but rather, “Does this system have a nontrivial solution?”

**Definition 4.1.4 (Solution space).** Let $A \in \mathbb{F}^{k \times n}$. The set of solutions to the homogeneous system of linear equations $Ax = 0$ is the set $U = \{x \in \mathbb{F}^n \mid Ax = 0\}$ and is called the *solution space* of $Ax = 0$. Ex. 4.1.7 explains the terminology.

**Definition 4.1.5 (Null space).** The *null space* or *kernel* of a matrix $A \in \mathbb{F}^{k \times n}$, denoted $\text{null}(A)$, is the set $\text{null}(A) = \{v \in \mathbb{F}^n \mid Av = 0\}$. (4.2)

**Definition 4.1.6.** The *nullity* of a matrix $A$ is the dimension of its null space.

For the following three exercises, let $A \in \mathbb{F}^{k \times n}$ and let $U$ be the solution space of the system $Ax = 0$.

**Exercise 4.1.7.** Prove that $U \subseteq \mathbb{F}^n$.

**Exercise 4.1.8.** Show that $\text{null}(A) = U$.

**Proposition 4.1.9.** Let $A \in \mathbb{F}^{k \times n}$ and consider the homogeneous system $Ax = b$.

(a) Let $\text{rk} A = r$ and let $n = r + d$. Then it is possible to relabel the unknowns $x_1, \ldots, x_n$ as $x'_1, \ldots, x'_n$, so that $x'_j$ can be represented as a linear combination of $x'_{d+1}, \ldots, x'_n$ for $j = 1, \ldots, d$, say

$$x'_i = \sum_{j=d+1}^{n} \lambda_{ij} x'_j.$$  

(b) The vectors $e_i + \sum_{j=d+1}^{n} \lambda_{ij} x'_j$ form a basis of $U'$, the solution space of the relabeled system of equations.

(c) $\dim U = \dim U'$

(d) $\dim U = n - \text{rk} A$

An immediate consequence of (d) is the Rank-Nullity Theorem, which will be crucial in our study of linear maps in Chapter 16.

**Corollary 4.1.10 (Rank-Nullity Theorem).** Let $A \in \mathbb{F}^{k \times n}$ be a matrix. Then

$$\text{rk}(A) + \text{nullity}(A) = n.$$  (4.3)

An explanation of (d) is that $\dim U$ measures the number of coordinates of $x$ that we can choose independently. This quantity is referred to by physicists as the “degree of freedom” left in our choice of $x$ after the set $Ax = b$ of constraints. If there are no constraints, the degree of freedom of the system is equal to $n$. It is plausible that each constraint reduces the degree of freedom by 1, which would suggest $\dim U = n - k$, but effectively there are only $\text{rk} A$ constraints because every equation that is

---

1In Chapter 16 we will formulate the Rank-Nullity Theorem in terms of linear maps rather than matrices, but the two formulations are equivalent.
a linear combination of previous equations can be thrown out. This makes it plausible that the degree of freedom is \( n - \text{rk} \, A \). This argument is not a proof, however.

**Proposition 4.1.11.** Let \( A \in \mathbb{F}^{k \times n} \) and consider the homogeneous system \( Ax = 0 \) of linear equations. Let \( U \) be the solution space of \( Ax = 0 \). Prove that the following are equivalent.

(a) \( Ax = 0 \) has no nontrivial solution

(b) The columns of \( A \) are linearly independent

(c) \( A \) has full column rank, i.e., \( \text{rk} \, A = n \)

(d) \( \dim U \neq 0 \)

(e) The rows of \( A \) span \( \mathbb{F}^n \)

(f) \( A \) has a left inverse

**Proposition 4.1.12.** Let \( A \) be a square matrix. Then \( Ax = 0 \) has no nontrivial solution if and only if \( A \) is nonsingular.

### 4.2 General systems of linear equations

**Proposition 4.2.1.** The system of linear equations \( Ax = b \) is solvable if and only if \( b \in \text{col-space} \, A \).

**Definition 4.2.2** (Augmented system). When speaking of the system \( Ax = b \), we call \( A \) the matrix of the system and \( [A \mid b] \) (the column \( b \) added to \( A \)) the augmented system.

**Proposition 4.2.3.** The system \( Ax = b \) of linear equations is solvable if and only if the matrix of the system and the augmented matrix have the same rank, i.e., \( \text{rk} \, A = \text{rk} \, [A \mid b] \).

**Definition 4.2.4** (Translation). Let \( S \subseteq \mathbb{F}^n \), and let \( v \in \mathbb{F}^n \). The set

\[
S + v = \{s + v \mid s \in S\}
\]

is called the translate of \( S \) by \( v \). Such an object is called an affine subspace of \( \mathbb{F}^n \) (\( \mathbb{F}^n \) Def. 5.1.3).

**Proposition 4.2.5.** Let \( S = \{x \in \mathbb{F}^n \mid Ax = b\} \) be the set of solutions to a system of linear equations. Then \( S \) is either empty or a translate of a subspace of \( \mathbb{F}^n \), namely, of the solution space of to the homogeneous system \( Ax = 0 \).

**Proposition 4.2.6** (Equivalent characterizations of nonsingular matrices). Let \( A \in M_n(\mathbb{F}) \). The following are equivalent.

(a) \( A \) is nonsingular, i.e., \( \text{rk} \, (A) = n \)

(b) The rows of \( A \) are linearly independent

(c) The columns of \( A \) are linearly independent

(d) The rows of \( A \) span \( \mathbb{F}^n \)

(e) The columns of \( A \) span \( \mathbb{F}^n \)
(f) The rows of $A$ form a basis of $\mathbb{F}^n$

(g) The columns of $A$ form a basis of $\mathbb{F}^n$

(h) $A$ has a left inverse

(i) $A$ has a right inverse

(j) $A$ has a two-sided inverse

(k) $Ax = 0$ has no nontrivial solution

(l) For all $b \in \mathbb{F}^n$, $Ax = b$ is solvable

(m) For all $b \in \mathbb{F}^n$, $Ax = b$ has a unique solution

Our discussion of the determinant in the next chapter will allow us to add a particularly important additional property:

(n) $\det A \neq 0$
Chapter 5

(𝔽, ℝ) Affine and Convex Combinations (optional)

The reader comfortable with abstract vector spaces can skip to Chapter ?? for a more general discussion of this material.

5.1 (𝔽) Affine combinations

In Section 1.1 we defined linear combinations of column vectors (𝔽 Def. [1.1.13]). We now consider affine combinations.

Definition 5.1.1 (Affine combination). An affine combination of the vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^n \) is a linear combination \( \sum_{i=1}^{k} \alpha_i \mathbf{v}_i \) where \( \sum_{i=1}^{k} \alpha_i = 1 \).

Example 5.1.2.

\[ 2\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2 - \frac{1}{3}\mathbf{v}_4 - \frac{1}{6}\mathbf{v}_5 \]

is an affine combination of the column vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_5 \).

Definition 5.1.3 (Affine-closed set). The set \( S \subseteq \mathbb{F}^n \) is affine-closed if it is closed under affine combinations.

Fact 5.1.4. The empty set is affine-closed (why?).

Definition 5.1.5 (Affine subspace). A nonempty, affine-closed subset \( U \) of \( \mathbb{F}^n \) is called an affine subspace of \( \mathbb{F}^n \) (notation: \( U \leq_{\text{aff}} \mathbb{F}^n \)).

Proposition 5.1.6. The intersection of a (finite or infinite) family of affine-closed subsets of \( \mathbb{F}^n \) is affine-closed. In other words, an intersection affine subspaces is either empty or an affine subspace.

Throughout this book, the term “subspace” refers to subsets that are closed under linear combinations (𝔽 Def. [1.2.1]). Subspaces are also referred to as “linear subspaces.” This (redundant) longer term is especially useful in contexts where affine subspaces are discussed in order to distinguish linear subspaces from affine subspaces.

Definition 5.1.7 (Affine hull). The affine hull of a subset \( S \subseteq \mathbb{F}^n \), denoted \( \text{aff } S \), is the smallest affine-closed set containing \( S \), i.e.,
(a) \( \text{aff } S \supseteq S \);

(b) \( \text{aff } S \) is affine-closed;

(c) for every affine set \( T \subseteq \mathbb{F}^n \), if \( T \supseteq S \) then \( T \supseteq \text{aff } S \).

**Fact 5.1.8.** \( \text{aff } \emptyset = \emptyset \).

Contrast this with the fact that \( \text{span } \emptyset = \{0\} \).

**Theorem 5.1.9.** Let \( S \subseteq \mathbb{F}^n \). Then \( \text{aff } S \) exists and is unique.

**Theorem 5.1.10.** For \( S \subseteq \mathbb{F}^n \), \( \text{aff } S \) is the set of all affine combinations of the finite subsets of \( S \).

**Proposition 5.1.11.** Let \( S \subseteq \mathbb{F}^n \). Then \( \text{aff } (\text{aff } S) = \text{aff } S \).

**Proposition 5.1.12.** Let \( S \subseteq \mathbb{F}^n \). Then \( S \) is affine-closed if and only if \( S = \text{aff } S \).

**Proposition 5.1.13.** Let \( S \subseteq \mathbb{F}^n \) be affine-closed. Then \( S \subseteq \mathbb{F}^n \) if and only if \( 0 \in S \).

**Proposition 5.1.14.**

(a) Let \( W \subseteq \mathbb{F}^n \). All translates \( W + v \) of \( W \) (Def. 4.2.4) are affine subspaces.

(b) Every affine subspace \( S \subseteq \mathbb{F}^n \) is the translate of a (unique) subspace of \( \mathbb{F}^n \).

**Proposition 5.1.15.** The intersection of a (finite or infinite) family of affine subspaces is either empty or equal to a translate of the intersection of their corresponding linear subspaces.

Next we connect these concepts with the theory of systems of linear equations.

**Exercise 5.1.16.** Let \( A \in \mathbb{F}^{k \times n} \) and let \( b \in \mathbb{F}^k \). Then the set of solutions to the system \( Ax = b \) of linear equations is an affine-closed subset of \( \mathbb{F}^n \).

The next exercise shows the converse.

**Exercise 5.1.17.** Every affine-closed subset of \( \mathbb{F}^n \) is the set of solutions to the system \( Ax = b \) of linear equations for some \( A \in \mathbb{F}^{k \times n} \) and \( b \in \mathbb{F}^k \).

**Proposition 5.1.18** (General vs. homogeneous systems of linear equations). Let \( A \in \mathbb{F}^{k \times n} \) and \( b \in \mathbb{F}^n \). Let \( S = \{ x \in \mathbb{F}^n \mid Ax = b \} \) be the set of solutions of the system \( Ax = b \) and let \( U = \{ x \in \mathbb{F}^n \mid Ax = 0 \} \) be the set of solutions of the corresponding system of homogeneous linear equations. Then either \( S \) is empty or \( S \) is a translate of \( U \).

We now study geometric features of \( \mathbb{F}^n \), viewed as an affine space.

**Proposition 5.1.19.** The span of the set \( S \subseteq \mathbb{F}^n \) is the affine hull of \( S \cup \{0\} \).

**Proposition 5.1.20.** Let \( S \subseteq W \) be nonempty. Then for any \( u \in S \), we have

\[
\text{aff } S = u + \text{span}(S - u)
\]

where \( S - u \) is the translate of \( S \) by \( -u \).
5.2. (F) HYPERPLANES

Definition 5.1.21 (Dimension of an affine subspace). The (affine) dimension of an affine subspace \( U \leq_{\text{aff}} \mathbb{F}^n \), denoted \( \dim_{\text{aff}} U \), is the dimension of its corresponding linear subspace (of which it is a translate). In order to assign a dimension to all affine-closed sets, we adopt the convention that \( \dim \emptyset = -1 \).

Exercise 5.1.22. Let \( U \leq \mathbb{F}^k \). Then \( \dim_{\text{aff}} U = \dim U \).

Exercise 5.1.23. What are the 0-dimensional affine subspaces?

Definition 5.1.24 (Affine independence). The vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^n \) are affine-independent if for every \( \alpha_1, \ldots, \alpha_k \in \mathbb{F} \), the two conditions
\[
\sum_{i=1}^{k} \alpha_i \mathbf{v}_i = 0 \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i = 0 \implies \alpha_1 = \cdots = \alpha_k = 0.
\]

Proposition 5.1.25. The vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^n \) are affine-independent if and only if none of them belongs to the affine hull of the others.

Fact 5.1.26. Any single vector is affine-independent and affine-closed at the same time.

Proposition 5.1.27 (Translation invariance). Let \( \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^n \) be affine-independent and let \( \mathbf{w} \in \mathbb{F}^n \) be any vector. Then \( \mathbf{v}_1 + \mathbf{w}, \ldots, \mathbf{v}_k + \mathbf{w} \) are also affine-independent.

Proposition 5.1.28 (Affine vs. linear independence). Let \( \mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^n \).
\begin{enumerate}
\item For \( k \geq 0 \), the vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \) are linearly independent if and only if the vectors \( 0, \mathbf{v}_1, \ldots, \mathbf{v}_k \) are affine-independent.
\item For \( k \geq 1 \), the vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \) are affine-independent if and only if the vectors \( \mathbf{v}_2 - \mathbf{v}_1, \ldots, \mathbf{v}_k - \mathbf{v}_1 \) are linearly independent.
\end{enumerate}

Definition 5.1.29 (Affine basis). An affine basis of an affine subspace \( W \leq_{\text{aff}} \mathbb{F}^n \) is an affine-independent set \( S \) such that \( \text{aff } S = W \).

Proposition 5.1.30. Let \( W \) be an affine subspace of \( \mathbb{F}^n \). Every affine basis of \( W \) has \( 1 + \dim W \) elements.

Corollary 5.1.31. If \( W_1, \ldots, W_k \) are affine subspaces of \( \mathbb{F}^n \), then
\[
\dim_{\text{aff}} (\text{aff}\{W_1, \ldots, W_k\}) \leq (k-1) + \sum_{i=1}^{k} \dim_{\text{aff}} W_i.
\]

5.2 (F) Hyperplanes

Definition 5.2.1 (Linear hyperplane). A linear hyperplane of \( \mathbb{F}^n \) is a subspace of \( \mathbb{F}^n \) of codimension 1 (F Def. 3.5.1).

Definition 5.2.2 (Codimension of an affine subspace). The (affine) codimension of an affine subspace \( U \leq_{\text{aff}} \mathbb{F}^n \), denoted \( \text{codim}_{\text{aff}} U \), is the codimension of its corresponding linear subspace (of which it is a translate).

Definition 5.2.3 (Hyperplane). A hyperplane is an affine subspace of codimension 1.
**Proposition 5.2.4.** Let $S \subseteq \mathbb{F}^n$ be a hyperplane. Then there exist a nonzero vector $a \in \mathbb{F}^n$ and $\beta \in \mathbb{F}$ such that $a^T v = \beta$ if and only if $v \in S$.

The vector $a$ whose existence is guaranteed by the preceding proposition is called the *normal vector* of the hyperplane $S$.

**Proposition 5.2.5.** Let $W \leq \mathbb{F}^n$. Then

(a) $W$ is the intersection of linear hyperplanes;

(b) if $\dim W = k$, then $W$ is the intersection of $n-k$ linear hyperplanes.

**Proposition 5.2.6.** Let $W \leq_{\text{aff}} \mathbb{F}^n$ be an affine subspace. Then

(a) $W$ is the intersection of hyperplanes;

(b) if $\dim_{\text{aff}} W = k$, then $W$ is the intersection of $n-k$ hyperplanes.

### 5.3 (R) Convex combinations

In the preceding section, we studied affine combinations over an arbitrary field $\mathbb{F}$. We now restrict ourselves to the case where $\mathbb{F} = \mathbb{R}$.

**Definition 5.3.1 (Convex combination).** A convex combination is an affine combination with nonnegative coefficients. So the expression

$$\sum_{i=1}^{k} \alpha_i v_i$$

is a convex combination of the vectors $v_1, \ldots, v_5$. Note that the affine combination in Example 5.1.2 is not convex.

**Example 5.3.2.**

$$\frac{1}{2} v_1 + \frac{1}{4} v_2 + \frac{1}{6} v_4 + \frac{1}{12} v_5$$

is a convex combination of the vectors $v_1, \ldots, v_5$.

**Fact 5.3.3.** Every convex combination is an affine combination.

**Definition 5.3.4 (Convex set).** A convex set is a subset $S \subseteq \mathbb{R}^n$ that is closed under convex combinations.

**Proposition 5.3.5.** The intersection of a (finite or infinite) family of convex sets is convex.

**Definition 5.3.6.** The convex hull of a subset $S \subseteq \mathbb{R}^n$, denoted conv $S$, is the smallest convex set containing $S$, i.e.,

(a) conv $S \supseteq S$;

(b) conv $S$ is convex;

(c) for every convex set $T \subseteq \mathbb{R}^n$, if $T \supseteq S$ then $T \supseteq$ conv $S$.

**Theorem 5.3.7.** Let $S \subseteq \mathbb{R}^n$. Then conv $S$ exists and is unique. \(\Diamond\)

**Theorem 5.3.8.** For $S \subseteq \mathbb{R}$, conv $S$ is the set of all convex combinations of the finite subsets of $S$. \(\Diamond\)

**Proposition 5.3.9.** Let $S \subseteq \mathbb{R}^n$. Then $S$ is convex if and only if $S = \text{conv } S$. 

5.4. \((\mathbb{R})\) HELLY’S THEOREM

Fact 5.3.10. \(\text{conv } S \subseteq \text{aff } S\).

Definition 5.3.11 (Straight-line segment). Let \(u, v \in \mathbb{R}^n\). The straight-line segment connecting \(u\) and \(v\) is the convex hull of \(\{u, v\}\), i.e., the set

\[
\text{conv}(u, v) = \{\lambda u + (1 - \lambda)v \mid 0 \leq \lambda \leq 1\}.
\]

Proposition 5.3.12. The set \(S \subseteq \mathbb{R}^n\) is convex if and only if it contains the straight-line segment connecting \(u\) and \(v\) for every \(u, v \in S\).

Definition 5.3.13 (Dimension). The dimension of a convex set \(C \subseteq \mathbb{R}^n\), denoted \(\dim_{\text{conv}} C\), is the dimension of its affine hull. A convex subset \(C\) of \(\mathbb{R}^n\) is full-dimensional if \(\text{aff } C = \mathbb{R}^n\).

Definition 5.3.14 (Half-space). A closed half-space is a region of \(\mathbb{R}^n\) defined as \(\{v \in \mathbb{R}^n \mid a^Tv \geq \beta\}\) for some nonzero \(a \in \mathbb{R}^n\) and \(\beta \in \mathbb{R}\) and is denoted \(H(a, \beta)\). An open half-space is a region of \(\mathbb{R}^n\) defined as \(\{v \in \mathbb{R}^n \mid a^Tv > \beta\}\) for some nonzero \(a \in \mathbb{R}^n\) and \(\beta \in \mathbb{R}\) and is denoted \(H^o(a, \beta)\).

Exercise 5.3.15. Let \(a \in \mathbb{R}^n\) be a nonzero vector and let \(\beta \in \mathbb{R}\). Prove: the set \(\{v \in \mathbb{R}^n \mid a^Tv \leq \beta\}\) is also a (closed) half-space.

Fact 5.3.16. Let \(S \subseteq \mathbb{R}^n\) be a hyperplane defined by \(a^Tv = \beta\). Then \(S\) divides \(\mathbb{R}^n\) into two half-spaces, defined by \(a^Tv \geq \beta\) and \(a^Tv \leq \beta\). The intersection of these two half-spaces is \(S\).

Proposition 5.3.17. (a) Every closed \((\mathbb{R})\text{Def. } 19.4.11\) convex set is the intersection of the closed half-spaces containing it.

(b) If \(S \subseteq \mathbb{R}^n\) is finite then \(\text{conv}(S)\) is the intersection of a finite number of closed half-spaces.

Exercise 5.3.18. Find a convex set which is not the intersection of any number of open or closed half-spaces. (Such a set already exists in 2 dimensions.)

Proposition* 5.3.19. Every open \((\mathbb{R})\text{Def. } 19.4.10\) convex set is the intersection of the open half-spaces containing it.

5.4 \((\mathbb{R})\) Helly’s Theorem

The main result of this section is the following theorem.

Theorem 5.4.1 (Helly’s Theorem). If \(C_1, \ldots, C_k \subseteq \mathbb{R}^n\) are convex sets such that any \(n+1\) of them intersect then all of them intersect.

Lemma 5.4.2 (Radon). Let \(S \subseteq \mathbb{R}^n\) be a set of \(k \geq n+2\) vectors in \(\mathbb{R}^n\). Then \(S\) has two disjoint subsets \(S_1\) and \(S_2\) whose convex hulls intersect.

Exercise 5.4.3. Show that the inequality \(k \geq n+2\) is tight, i.e., that there exists a set \(S\) of \(n+1\) vectors in \(\mathbb{R}^n\) such that for any two subsets \(S_1, S_2 \subseteq S\), we have \(\text{conv } S_1 \cap \text{conv } S_2 = \emptyset\).

Exercise 5.4.4. Use Radon’s Lemma to prove Helly’s Theorem.
Exercise 5.4.5. Show the bound in Helly's Theorem is tight, i.e., there exist $n + 1$ convex subsets of $\mathbb{R}^n$ such that every $n$ of them intersect but the intersection of all of them is empty.

5.5 $\langle \mathbb{F}, \mathbb{R} \rangle$ Additional exercises

Observe that a half-space can be defined by $n + 1$ real parameters, i.e., by defining $a$ and $\beta$. The following project asks you to define a similar object which can be defined by $2n - 1$ real parameters.

Project 5.5.1. Define a “partially open half-space” which can be defined by $2n - 1$ parameters so that every convex set is the intersection of the partially open half-spaces and the closed half-spaces containing it.

---

1 In fact, in a well-defined sense, half-spaces can be defined by $n$ real parameters.
2 In fact, this object can be defined by $2n - 3$ real parameters.
Chapter 6

(\text{F}) The Determinant

While many systems of linear equations are too large to reasonably solve by hand using methods like elimination, the general system of two equations in two unknowns

\[ \begin{align*}
\alpha_{11}x_1 + \alpha_{12}x_2 &= \beta_1 \quad (6.1) \\
\alpha_{21}x_1 + \alpha_{22}x_2 &= \beta_2 \quad (6.2)
\end{align*} \]

is not too cumbersome to work out and find

\[ \begin{align*}
x_1 &= \frac{\alpha_{22}\beta_1 - \alpha_{12}\beta_2}{\alpha_{22}\alpha_{11} - \alpha_{12}\alpha_{21}} \quad (6.3) \\
x_2 &= \frac{\alpha_{11}\beta_2 - \alpha_{21}\beta_1}{\alpha_{22}\alpha_{11} - \alpha_{12}\alpha_{21}} \quad (6.4)
\end{align*} \]

For the case of 3 equations in 3 unknowns, we obtain

\[ \begin{align*}
x_1 &= \frac{N_{31}}{D_3} \quad (6.5) \\
x_2 &= \frac{N_{32}}{D_3} \quad (6.6) \\
x_3 &= \frac{N_{33}}{D_3} \quad (6.7)
\end{align*} \]

where

\[ \begin{align*}
D_3 &= \alpha_{11}\alpha_{22}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} \\
&\quad - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} - \alpha_{13}\alpha_{22}\alpha_{31} \\
&\quad - \alpha_{13}\alpha_{23}\alpha_{31} - \alpha_{23}\alpha_{11}\alpha_{32} - \alpha_{22}\alpha_{12}\alpha_{31} \\
&\quad - \alpha_{21}\alpha_{33}\alpha_{12} - \alpha_{13}\alpha_{31}\alpha_{22} - \alpha_{12}\alpha_{31}\alpha_{23} \\
&\quad - \alpha_{11}\alpha_{32}\alpha_{23} - \alpha_{23}\alpha_{12}\alpha_{31} - \alpha_{21}\alpha_{31}\alpha_{12} \\
&\quad - \alpha_{21}\alpha_{33}\alpha_{12} - \alpha_{23}\alpha_{11}\alpha_{32} - \alpha_{22}\alpha_{12}\alpha_{31}
\end{align*} \]

\[ \begin{align*}
N_{31} &= \alpha_{22}\alpha_{33}\beta_1 + \alpha_{12}\alpha_{23}\beta_3 + \alpha_{13}\alpha_{32}\beta_2 \\
&\quad - \alpha_{23}\alpha_{32}\beta_1 - \alpha_{12}\alpha_{33}\beta_2 - \alpha_{13}\alpha_{22}\beta_3 \\
N_{32} &= \alpha_{11}\alpha_{33}\beta_2 + \alpha_{23}\alpha_{31}\beta_1 + \alpha_{13}\alpha_{32}\beta_3 \\
&\quad - \alpha_{11}\alpha_{23}\beta_3 - \alpha_{21}\alpha_{33}\beta_1 - \alpha_{13}\alpha_{31}\beta_2 \\
N_{33} &= \alpha_{11}\alpha_{22}\beta_3 + \alpha_{12}\alpha_{31}\beta_2 + \alpha_{21}\alpha_{32}\beta_1 \\
&\quad - \alpha_{11}\alpha_{32}\beta_2 - \alpha_{12}\alpha_{21}\beta_3 - \alpha_{22}\alpha_{31}\beta_1
\end{align*} \]

In particular, the numerators and denominators of these expressions each have six terms. For a system of 4 equations in 4 unknowns, the numerators and denominators each have 24 terms, and, in general, we have $n!$ terms in the numerators and denominators of the solutions to $n$ equations in $n$ unknowns.
As complicated as these expressions may seem for large \( n \), it is clear that they represent a fundamental quantity associated with the matrix.

The denominator of these expressions is known as the determinant of the matrix \( A \) (where our system of linear equations is written as \( Ax = b \)).

The determinant is a function from the space of \( n \times n \) matrices to numbers, that is, \( \det : M_n(\mathbb{F}) \to \mathbb{F} \). Before defining this function, we need to study permutations.

### 6.1 Permutations

**Definition 6.1.1 (Permutation).** A permutation of a set \( \Omega \) is a bijection \( f : \Omega \to \Omega \). The set \( \Omega \) is called the permutation domain.

**Definition 6.1.2 (Symmetric group).** The symmetric group of degree \( n \), denoted \( S_n \), is the set of all permutations of the set \( \{1, \ldots, n\} \).

**Definition 6.1.3 (Inversion).** Let \( \sigma \in S_n \) be a permutation of the set \( \{1, \ldots, n\} \), and let \( 1 \leq i, j \leq n \) with \( i \neq j \). We say that the pair \( \{i, j\} \) is inverted by \( \sigma \) if \( i < j \) and \( \sigma(i) > \sigma(j) \) or \( i > j \) and \( \sigma(i) < \sigma(j) \). We denote by \( \text{Inv}(\sigma) \) the number of inversions of \( \sigma \), that is, the number of pairs \( \{i, j\} \) that are inverted by \( \sigma \).

**Exercise 6.1.4.** What is the maximum possible value of \( \text{Inv}(\sigma) \) for \( \sigma \in S_n \)?

**Definition 6.1.5 (Even and odd permutations).** If \( \text{Inv}(\sigma) \) is even, then we say that \( \sigma \) is an even permutation, and if \( \text{Inv}(\sigma) \) is odd, then \( \sigma \) is an odd permutation.

**Proposition 6.1.6.** Half of the \( n! \) permutations of \( \{1, \ldots, n\} \) are even.

**Definition 6.1.7 (Sign of a permutation).** Let \( \sigma \in S_n \) be a permutation. The sign of \( \sigma \), denoted \( \text{sgn} \sigma \), is defined as

\[
\text{sgn}(\sigma) := (-1)^{\text{Inv}(\sigma)}
\]

In particular, \( \text{sgn}(\sigma) = 1 \) if \( \sigma \) is an even permutation and \( \text{sgn}(\sigma) = -1 \) if \( \sigma \) is an odd permutation. The sign of a permutation (equivalently, whether it is even or odd) is also referred to as the parity of the permutation.

**Definition 6.1.8 (Composition of permutations).** Let \( \Omega \) be a set, and let \( \sigma \) and \( \tau \) be permutations of \( \Omega \). Then the composition of \( \sigma \) with \( \tau \), denoted \( \sigma \tau \), is defined by

\[
(\sigma \tau)(a) := \sigma(\tau(a))
\]

for all \( a \in \Omega \).

We also refer to the composition of \( \sigma \) with \( \tau \) as the product of \( \sigma \) and \( \tau \).

**Definition 6.1.9 (Transposition).** Let \( \Omega \) be a set. The transposition of the elements \( a \neq b \in \Omega \) is the permutation that swaps \( a \) and \( b \) and fixes every other element. Formally, it is the permutation \( \tau \) defined by

\[
\tau(x) := \begin{cases} 
  b & x = a \\
  a & x = b \\
  x & \text{otherwise}
\end{cases}
\]

This permutation is denoted \( \tau = (a, b) \).
Proposition 6.1.10. Transpositions generate $S_n$, i.e., every permutation $\sigma \in S_n$ can be written as a composition of transpositions.

Theorem 6.1.11. A permutation $\sigma$ of $\{1, \ldots, n\}$ is even if and only if $\sigma$ is a product of an even number of transpositions.

In the light of this theorem, we could use its conclusion as the definition of even permutations. The advantage of this definition is that it can be applied to any set $\Omega$, not just the ordered set $\{1, \ldots, n\}$.

Corollary 6.1.12. While the number of inversions of a permutation depends on the ordering of the permutation domain, its parity does not.

Definition 6.1.13 (Neighbor transposition). A neighbor transposition of the set $\{1, \ldots, n\}$ is a transposition of the form $\tau = (i, i+1)$.

Exercise 6.1.14. Let $\sigma \in S_n$ and let $\tau$ be a neighbor transposition. Show

$$|\text{Inv}(\sigma) - \text{Inv}(\sigma\tau)| = 1.$$ 

Corollary 6.1.15. Let $\sigma \in S_n$ and let $\tau$ be a neighbor transposition. Then $\text{sgn}(\sigma\tau) = -\text{sgn}(\sigma)$.

Proposition 6.1.16. Every transposition is the composition of an odd number of neighbor transpositions.

Proposition 6.1.17. Neighbor transpositions generate the symmetric group. That is, every element of $S_n$ can be expressed as the composition of neighbor transpositions.

Proposition 6.1.18. Composition with a transposition changes the parity of a permutation.

Corollary 6.1.19. Let $\sigma \in S_n$ be a permutation. Then $\sigma$ is even if and only if $\sigma$ is the product of an even number of transpositions.

Theorem 6.1.20. Let $\sigma, \tau \in S_n$. Then

$$\text{sgn}(\sigma\tau) = \text{sgn}(\sigma) \text{sgn}(\tau).$$

Definition 6.1.21 (k-cycle). A k-cycle is a permutation that cyclically permutes k elements $\{a_1, \ldots, a_k\}$ and fixes all others (FIGURE). That is, $\sigma$ is a k-cycle if $\sigma(a_i) = a_{i+1}$ for some elements $a_1, \ldots, a_k$ (where $a_{k+1} = a_1$) and $\sigma(x) = x$ if $x \notin \{a_1, \ldots, a_k\}$.

In particular, transpositions are 2-cycles.

Definition 6.1.22 (Disjoint cycles). Let $\sigma$ and $\tau$ be cycles with permutation domain $\Omega$. Then $\sigma$ and $\tau$ are disjoint if no element of $\Omega$ is permuted by both $\sigma$ and $\tau$.

Proposition 6.1.23. Every permutation uniquely decomposes into disjoint cycles.

Exercise 6.1.24. Let $\sigma$ be a k-cycle. Show that $\sigma$ is an even permutation if and only if k is odd.

Corollary 6.1.25. Let $\sigma$ be a permutation. Then $\sigma$ is even if and only if its cycle decomposition includes an even number of even cycles.

Proposition 6.1.26. Let $\sigma$ be a permutation. Then $\text{Inv}(\sigma^{-1}) = \text{Inv}(\sigma)$.
6.2 Defining the determinant

**Definition 6.2.1 (Determinant).** The determinant of an \( n \times n \) matrix \( A = (\alpha_{ij}) \) is

\[
\det A := \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} \alpha_{i,\sigma(i)}. \tag{6.20}
\]

**Notation 6.2.2.** The determinant of a matrix

\[
A = \begin{pmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{n1} & \cdots & \alpha_{nn}
\end{pmatrix}
\]

may also be represented with vertical bars, that is,

\[
\det A = |A| = \det \begin{pmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{n1} & \cdots & \alpha_{nn}
\end{pmatrix} \tag{6.21}
\]

or

\[
\begin{vmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{n1} & \cdots & \alpha_{nn}
\end{vmatrix}. \tag{6.22}
\]

**Proposition 6.2.3.** Show

(a) \( \det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \alpha \delta - \beta \gamma \)

(b) \( \det \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} = \alpha_{11}\alpha_{22}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} - \alpha_{13}\alpha_{22}\alpha_{31} \)

For the following exercises, let \( A \) be an \( n \times n \) matrix.

**Proposition 6.2.4.** If a column of \( A \) is 0, then \( \det A = 0 \).

**Proposition 6.2.5.** Let \( A' \) be the matrix obtained by swapping two columns of \( A \). Then \( \det A' = -\det A \).

**Proposition 6.2.6.** Show that \( \det A^T = \det A \). This fact follows from what property of inversions?

**Proposition 6.2.7.** If two columns of \( A \) are equal, then \( \det A = 0 \).

**Proposition 6.2.8.** The determinant of a diagonal matrix is the product of its diagonal entries.

**Proposition 6.2.9.** The determinant of an upper triangular matrix is the product of its diagonal entries.

**Example 6.2.10.**

\[
\begin{vmatrix}
5 & 1 & 7 \\
0 & 2 & 6 \\
0 & 0 & 3
\end{vmatrix} = 30 \tag{6.23}
\]

and this value does not depend on the three entries in the upper-right corner.

**Definition 6.2.11 (k-linearity).** A function \( f : V \times \cdots \times V \rightarrow W \) is linear in the \( i \)-th component, if, whenever we fix \( x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k \), the function

\[
g(y) := f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_k)
\]
6.2. DEFINING THE DETERMINANT

is a linear, i.e.,

\[ g(y_1 + y_2) = g(y_1) + g(y_2) \]  
\[ g(\alpha y) = \alpha g(y) \]  \hspace{1cm} (6.24)
\[ g(\alpha y) = \alpha g(y) \]  \hspace{1cm} (6.25)

The function \( f \) is \([k\text{-linear}]\) if it is linear in all \( k \) components. A function which is 2-linear is said to be \textit{bilinear}.

**Proposition 6.2.12** (Multilinearity). The determinant is multilinear in the columns of \( A \). That is,

\[
\begin{vmatrix}
    a_1 & \cdots & a_i - 1 & a_i + b & a_{i+1} & \cdots & a_n \\
\end{vmatrix} = 
\begin{vmatrix}
    a_1 & \cdots & a_n \\
\end{vmatrix} + 
\begin{vmatrix}
    a_1 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_n \\
\end{vmatrix}
\]  \hspace{1cm} (6.26)

and

\[
\begin{vmatrix}
    a_1 & \cdots & a_i - 1 & \alpha a_i & a_{i+1} & \cdots & a_n \\
\end{vmatrix} = \alpha \begin{vmatrix}
    a_1 & \cdots & a_n \\
\end{vmatrix} \hspace{1cm} (6.27)
\]

Next we study the effect of elementary column and row operations (Sec. 3.2) on the determinant.

**Proposition 6.2.13.** Let \( A' \) be the matrix obtained from \( A \in M_n(\mathbb{F}) \) by applying an elementary column or row operation. We distinguish cases according to the operation applied.

(a) Shearing does not change the value of the determinant: \( \det(A') = \det(A) \).

(b) Scaling of a row or column by a scalar \( \lambda \) (the scale_\text{e}(i, \lambda) or scale_\text{r}(i, \lambda) operation scales the value of the determinant by the same factor: \( \det(A') = \lambda \det(A) \).

(c) Swapping two rows or columns changes the sign of the determinant: \( \det(A') = -\det(A) \).

**Proposition 6.2.14.** Let \( A, B \in M_n(\mathbb{F}) \). Then \( \det(AB) = \det(A) \det(B) \).

**Numerical exercise 6.2.15.** Let \( A = \begin{pmatrix} 2 & 1 & -3 \\ 4 & -1 & 0 \\ 2 & 5 & -1 \end{pmatrix} \).

(a) Use the formula derived in part \( \text{(b)} \) of Prop. 6.2.3 to compute \( \det A \).

(b) Compute the matrix \( A' \) obtained by performing the column operation \((1, 2, -4)\) on \( A \).

(c) \textit{Self-check}: Use the same formula to compute \( \det A' \), and verify that \( \det A' = \det A \).

**Theorem 6.2.16** (Determinant vs. linear independence). Let \( A \) be a square matrix. Then \( \det A = 0 \) if and only if the columns of \( A \) are linearly dependent.

The proof of this theorem requires the following two lemmas which also provide a practical method to compute the determinant by Gaussian or Gauss-Jordan elimination.

\(^{1}\)Linear maps will be discussed in more detail in Chapter 16.
Lemma 6.2.17. Let $A$ be an $n \times n$ matrix whose columns are linearly dependent. Then through a series of column shearing operations, it is possible to bring $A$ into a form in which one column is all zeros. In particular, in this case the determinant is zero. ♦

Lemma 6.2.18. Let $A$ be an $n \times n$ matrix whose columns are linearly independent. Then it is possible to perform a series of column shearing operations that bring $A$ into a form in which each row and column contains exactly one nonzero entry, i.e., the determinant has exactly one nonzero expansion term. ♦

Theorem 6.2.19 (Determinant vs. system of linear equations). The homogeneous system $Ax = 0$ of $n$ linear equations in $n$ unknowns has a nontrivial solution if and only if $\det A = 0$. ♦

Definition 6.2.20 (Skew-symmetric matrix). The matrix $A \in M_n(\mathbb{F})$ is skew-symmetric if $A^T = -A$.

Exercise 6.2.21. Let $\mathbb{F}$ be a field in which $1 + 1 \neq 0$.

(a) Show that if $A \in M_n(\mathbb{F})$ is skew-symmetric and $n$ is odd then $\det A = 0$.

(b) For all even $n$, find a nonsingular skew-symmetric matrix $A \in M_n(\mathbb{F})$.

Exercise 6.2.22. Show that part (a) of the preceding exercise is false over $\mathbb{F}_2$ (and every field of characteristic 2).

Exercise 6.2.23. Let $\mathbb{F}$ be a field of characteristic 2 and let $n$ be odd. If $A \in M_n(\mathbb{F})$ is symmetric\footnote{Observe that in a field of characteristic 2, “skew-symmetric” means the same thing as “symmetric.”} and all of its diagonal entries are 0, then $A$ is singular.

Proposition 6.2.24 (Hadamard’s Inequality). Let $A = [a_1 | \cdots | a_n] \in M_n(\mathbb{R})$. Prove

$$|\det A| \leq \prod_{j=1}^{n} ||a_j||.$$  \hspace{1cm} (6.28)

where $||a_j||$ is the norm of the vector $a_j$ (Def. 1.5.2).

Exercise 6.2.25. When does equality hold in Hadamard’s Inequality?

Definition 6.2.26 (Parallelepiped). Let $v_1, v_2 \in \mathbb{R}^2$. The parallelogram spanned by $v_1$ and $v_2$ is the set (PICTURE)

$$\text{parallelogram}(v_1, v_2) := \{\alpha v_1 + \beta v_2 | 0 \leq \alpha, \beta \leq 1\}.$$  \hspace{1cm} (6.29)

More generally, the parallelepiped spanned by the vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ is the set

$$\text{parallelepiped}(v_1, \ldots, v_n) := \left\{ \sum_{i=1}^{n} \alpha_i v_i | 0 \leq \alpha_i \leq 1 \right\}.$$  \hspace{1cm} (6.30)
6.3. COFACTOR EXPANSION

Exercise 6.2.27. Let \( a, b \in \mathbb{R}^2 \). Show that the area of the parallelogram spanned by these vectors (PICTURE) is \(| \det A |\), where \( A = [a | b] \in M_2(\mathbb{R})\).

The following theorem is a generalization of Ex. 6.2.27.

Theorem 6.2.28. If \( v_1, \ldots, v_n \in \mathbb{R}^n \), then the volume of parallelepiped\( (v_1, \ldots, v_n)\) is \(| \det A |\), where \( A \) is the \( n \times n \) matrix whose \( i \)-th column is \( v_i \).

\[ \Box \]

6.3 Cofactor expansion

Definition 6.3.1 (Cofactor). Let \( A \in M_n(\mathbb{F}) \) be a matrix. The \((i,j)\) cofactor of \( A \) is

\[ (-1)^{i+j} \det (\hat{A}_{ij}), \]

where \( \hat{A}_{ij} \) is the \((n-1) \times (n-1)\) matrix obtained by removing the \( i \)-th row and \( j \)-th column of \( A \).

Theorem 6.3.2 (Cofactor expansion). Let \( A = (\alpha_{ij}) \) be an \( n \times n \) matrix, and let \( C_{ij} \) be the \((i,j)\) cofactor of \( A \). Then for all \( i \),

\[ \det A = \sum_{j=1}^{n} \alpha_{ij} C_{ij}. \quad (6.31) \]

This is the cofactor expansion of \( A \) along the \( i \)-th row. Similarly, for all \( j \),

\[ \det A = \sum_{i=1}^{n} \alpha_{ij} C_{ij}. \quad (6.32) \]

This is the cofactor expansion of \( A \) along the \( j \)-th column. \( \Box \)

Numerical exercise 6.3.3. Compute the determinants of the following matrices by cofactor expansion (a) along the first row and (b) along the second column. Self-check: your answers should be the same.

\[ (a) \begin{pmatrix} 2 & 3 & 1 \\ 0 & -4 & -1 \\ 1 & -3 & 4 \end{pmatrix} \]

\[ (b) \begin{pmatrix} 3 & -3 & 2 \\ 4 & 7 & -1 \\ 6 & -4 & 2 \end{pmatrix} \]

\[ (c) \begin{pmatrix} 1 & 3 & 2 \\ 3 & -1 & 0 \\ 0 & 6 & 5 \end{pmatrix} \]

\[ (d) \begin{pmatrix} 6 & 2 & -1 \\ 0 & 4 & 1 \\ -3 & 1 & 1 \end{pmatrix} \]

Exercise 6.3.4. Compute the determinants of the following matrices.

\[ (a) \begin{pmatrix} \alpha & \beta & \beta & \cdots & \beta & \beta \\ \beta & \alpha & \beta & \cdots & \beta & \beta \\ \beta & \beta & \alpha & \cdots & \beta & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta & \beta & \beta & \cdots & \alpha & \beta \\ \beta & \beta & \beta & \cdots & \beta & \alpha \end{pmatrix} \]

\[ (b) \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \end{pmatrix} \]
CHAPTER 6. \( \mathbb{F} \) THE DETERMINANT

Exercise 6.3.5. Compute the determinant of the Vandermonde matrix \( (\text{Def. 2.5.9 generated by } a_1, \ldots, a_n) \).  

Definition 6.3.6 (Fixed point). Let \( \Omega \) be a set, and let \( f : \Omega \to \Omega \) be a permutation of \( \Omega \). We say that \( x \in \Omega \) is a fixed point of \( f \) if \( f(x) = x \). We say that \( f \) is fixed-point-free, or a derangement, if \( f \) has no fixed points.

\( \heartsuit \) Exercise 6.3.7. Let \( F_n \) denote the set of fixed-point free permutations of the set \( \{1, \ldots, n\} \). Decide, for each \( n \), whether the majority of \( F_n \) is odd or even. (The answer will depend on \( n \).)

6.4 Matrix inverses and the determinant

We now derive an explicit form for the inverse of an \( n \times n \) matrix. Our tool for this will be the cofactor expansion in addition to the skew cofactor expansion. In Theorem [6.3.2](#) we took the dot product of the \( i \)-th column of a matrix with the cofactors of corresponding to the \( i \)-th column. We now instead take the dot product of the \( i \)-th column with the cofactors corresponding to the \( j \)-th column.

Proposition 6.4.1 (Skew cofactor expansion). Let \( A = (\alpha_{ij}) \) be an \( n \times n \) matrix, and fix \( i \) and \( j \) (\( 1 \leq i, j \leq n \)) such that \( i \neq j \). Let \( C_{kj} \) be the \((k, j)\) cofactor of \( A \). Then

\[
\sum_{k=1}^{n} \alpha_{ki} C_{kj} = 0.
\]

Definition 6.4.2 (Adjugate of a matrix). Let \( A \in M_n(\mathbb{F}) \). Then the adjugate of \( A \), denoted \( \text{adj}(A) \), is the matrix whose \((i, j)\) entry is the \((j, i)\) cofactor of \( A \).

Theorem 6.4.3 (Explicit form of the matrix inverse). Let \( A \) be a nonsingular \( n \times n \) matrix. Then

\[
A^{-1} = \frac{1}{\det A} \text{adj}(A).
\]

\( \diamond \) Numerical exercise 6.4.4. Compute the inverses of the following matrices. **Self-check:** multiply your answer by the original matrix to get the identity.

\[
\begin{align*}
(a) & \quad \begin{pmatrix} 1 & -3 \\ -2 & 4 \end{pmatrix} \\
(b) & \quad \begin{pmatrix} -4 & 2 \\ -1 & -1 \end{pmatrix} \\
(c) & \quad \begin{pmatrix} 3 & 4 & -7 \\ -2 & 1 & -4 \\ 0 & -2 & 5 \end{pmatrix} \\
(d) & \quad \begin{pmatrix} -1 & 4 & 2 \\ -3 & 2 & -3 \\ 1 & 0 & 2 \end{pmatrix}
\end{align*}
\]
Proposition 6.4.5. If $A \in M_n(\mathbb{F})$ is invertible, then $\det A^{-1} = \frac{1}{\det A}$.

Exercise 6.4.6. When is $A \in M_n(\mathbb{Z})$ invertible over $\mathbb{Z}$? Give a simple condition in terms of $\det A$.

Exercise 6.4.7. Let $n$ be odd and let $A \in M_n(\mathbb{Z})$ be an invertible symmetric matrix whose diagonal entries are all 0. Show that $A^{-1} \notin M_n(\mathbb{Z})$.

6.5 Additional exercises

Notation 6.5.1. Let $A \in \mathbb{F}^{k \times n}$ and $B \in \mathbb{F}^{n \times k}$ be matrices, and let $I \subseteq [n]$. The matrix $A_I$ is the matrix whose columns are the columns of $A$ which correspond to the elements of $I$. The matrix $I_B$ is $(B^T_I)^T$, i.e., the matrix whose rows are the rows of $B$ which correspond to the elements of $I$.

Example 6.5.2. Let $A = \begin{pmatrix} 3 & 1 & 7 \\ 2 & 3 & 2 \end{pmatrix}$, $B = \begin{pmatrix} -4 & 1 \\ 6 & 2 \\ -1 & 0 \end{pmatrix}$, and let $I = \{1, 3\}$. Then $A_I = \begin{pmatrix} 3 & 7 \\ 2 & 2 \end{pmatrix}$ and $I_B = \begin{pmatrix} -4 & 1 \\ -1 & 0 \end{pmatrix}$.

Exercise 6.5.3 (Cauchy-Binet formula). Let $A \in \mathbb{F}^{k \times n}$ and let $B \in \mathbb{F}^{n \times k}$. Show that

$$\det(AB) = \sum_{I \subseteq [n]} \det(A_I) \det(I_B). \quad (6.35)$$
Chapter 7

(F) Theory of Systems of Linear Equations II: Cramer’s Rule

7.1 Cramer’s Rule

With the determinant in hand, we are now able to return to the problem of solving systems of linear equations. In particular, we have the following result.

Theorem 7.1.1. Let $A$ be a nonsingular square matrix. Then the system $Ax = b$ of $n$ linear equations in $n$ unknowns is solvable. ♦

In particular, if $A$ is a square matrix, then $Ax = b$ is solvable (and in fact has a unique solution) precisely when $A$ is invertible. It is easy to see that in this case, the solution to $Ax = b$ is given by $x = A^{-1}b$, but we now demonstrate that it is possible to determine the solution to this system of equations without ever computing a matrix inverse.

Theorem 7.1.2 (Cramer’s Rule). Let $A$ be an invertible $n \times n$ matrix over $\mathbb{F}$, and let $b \in \mathbb{F}^n$.

Let $a = A^{-1}b = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$. Then

(a) $Aa = b$

(b) For each $i \leq n$,

$$\alpha_i = \frac{\det A_i}{\det A}$$

(7.1)

where $A_i$ is the matrix obtained by replacing the $i$-th column of $A$ by $b$.

Numerical exercise 7.1.3. Use Cramer’s Rule to solve the following systems of linear equations. Self-check: plug your answers back into the original equations.

(a)

\begin{align*}
x_1 + 2x_2 &= 3 \\
x_1 - x_2 &= 6
\end{align*}

(b)

\begin{align*}
-x_1 + x_2 - x_3 &= 4 \\
2x_1 + 3x_2 + 4x_3 &= -2 \\
x_1 - x_2 - 3x_3 &= 3
\end{align*}
Chapter 8

(\(F\)) Eigenvectors and Eigenvalues

8.1 Eigenvector and eigenvalue basics

Definition 8.1.1 (Eigenvector). Let \(A \in M_n(F)\) and let \(v \in F^n\). Then \(v\) is an eigenvector of \(A\) if \(v \neq 0\) and there exists \(\lambda \in F\) such that \(Av = \lambda v\).

Definition 8.1.2 (Eigenvalue). Let \(A \in M_n(F)\). Then \(\lambda \in F\) is an eigenvalue of \(A\) if there exists a nonzero vector \(v \in F^n\) such that \(Av = \lambda v\).

Exercise 8.1.3. What are the eigenvalues and eigenvectors of the \(n \times n\) identity matrix?

Exercise 8.1.4. Let \(D = \text{diag}(\lambda_1, \ldots, \lambda_n)\) be the \(n \times n\) diagonal matrix whose diagonal entries are \(\lambda_1, \ldots, \lambda_n\). What are the eigenvalues and eigenvectors of \(D\)?

Exercise 8.1.5. Determine the eigenvalues of the \(2 \times 2\) matrix \(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\).

Exercise 8.1.6. Find examples of real \(2 \times 2\) matrices with

(a) no eigenvectors in \(\mathbb{R}^2\)
(b) exactly one eigenvector in \(\mathbb{R}^2\), up to scaling
(c) two eigenvectors in \(\mathbb{R}^2\), up to scaling
(d) infinitely many eigenvectors in \(\mathbb{R}^2\), up to scaling

\(\heartsuit\) Exercise 8.1.7. Show that eigenvectors of a matrix corresponding to distinct eigenvalues are linearly independent.

Proposition 8.1.8. Let \(A \in M_n(F)\) be a matrix and let \(v_1\) and \(v_2\) be eigenvectors to distinct eigenvalues. Then \(v_1 + v_2\) is not an eigenvector.

Proposition 8.1.9. Let \(A \in M_n(F)\) be a matrix to which every nonzero vector is an eigenvector. Then \(A\) is a scalar matrix (\(\mathbb{F}\) Def. 2.2.13).

Definition 8.1.10 (Eigenbasis). Let \(A \in M_n(F)\) be a matrix. An eigenbasis of \(A\) is a basis of \(F^n\) made up of eigenvectors of \(A\).

Exercise 8.1.11. Find all eigenbases of \(I\).

Exercise 8.1.12. Let \(\lambda_1, \ldots, \lambda_n\) be scalars. Find an eigenbasis of \(\text{diag}(\lambda_1, \ldots, \lambda_n)\).
Exercise 8.1.13. Prove that \( \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \) has no eigenbasis.

Proposition 8.1.14. Let \( A \in M_n(\mathbb{F}) \). Then \( A \) is nonsingular if and only if \( 0 \) is not an eigenvalue of \( A \).

Definition 8.1.15 (Left eigenvector). Let \( A \in M_n(\mathbb{F}) \). Then \( \mathbf{x} \in \mathbb{F}^{1 \times n} \) is a left eigenvector if \( \mathbf{x} \neq 0 \) and there exists \( \lambda \in \mathbb{F} \) such that \( \mathbf{x} A = \lambda \mathbf{x} \).

Definition 8.1.16 (Left eigenvalue). Let \( A \in M_n(\mathbb{F}) \). Then \( \lambda \in \mathbb{F} \) is a left eigenvalue of \( A \) if there exists a nonzero row vector \( \mathbf{v} \in \mathbb{F}^{1 \times n} \) such that \( \mathbf{v} A = \lambda \mathbf{v} \).

Convention 8.1.17. When we use the word “eigenvector” without a modifier, we refer to a right eigenvector; the term “right eigenvector” is occasionally used for clarity.

Exercise 8.1.18. Let \( A \in M_n(\mathbb{F}) \). Show that if \( \mathbf{x} \) is a right eigenvector to eigenvalue \( \lambda \) and \( \mathbf{y}^T \) is a left eigenvector to eigenvalue \( \mu \neq \lambda \), then \( \mathbf{y}^T \cdot \mathbf{x} = 0 \), i.e., \( \mathbf{y} \perp \mathbf{x} \).

Definition 8.1.19 (Eigenspace). Let \( A \in M_n(\mathbb{F}) \). We denote by \( U_\lambda \) the set
\[
U_\lambda := \{ \mathbf{v} \in \mathbb{F}^n \mid A \mathbf{v} = \lambda \mathbf{v} \}.
\]
This set is called the eigenspace corresponding to the eigenvalue \( \lambda \).

The following exercise explains this terminology.

Exercise 8.1.20. Let \( A \in M_n(\mathbb{F}) \) be a square matrix with eigenvalue \( \lambda \). Show that \( U_\lambda \) is a subspace of \( \mathbb{F}^n \).

Definition 8.1.21 (Geometric multiplicity). Let \( A \in M_n(\mathbb{F}) \) be a square matrix and let \( \lambda \) be an eigenvalue of \( A \). Then the geometric multiplicity of \( \lambda \) is \( \dim U_\lambda \).

The next exercise provides a method of calculating this quantity.

Exercise 8.1.22. Let \( \lambda \) be an eigenvalue of the \( n \times n \) matrix \( A \). Give a simple expression for the geometric multiplicity in terms of \( A \) and \( \lambda \).

Exercise 8.1.23. If \( N \) is a nilpotent matrix (Def. 2.2.18) and \( \lambda \) is an eigenvalue of \( N \), then \( \lambda = 0 \).

Exercise 8.1.24. Let \( N \) be a nilpotent matrix.

(a) Show that \( I + N \) is invertible.

(b) Find a nonsingular matrix \( A \) and a nilpotent matrix \( N \) such that \( A + N \) is singular.

(c) Prove that there exist a nonsingular diagonal matrix \( D \) and a nilpotent matrix \( N \) such that \( D + N \) is singular.

8.2 Similar matrices and diagonalizability

Definition 8.2.1 (Similar matrices). Let \( A, B \in M_n(\mathbb{F}) \). Then \( A \) and \( B \) are similar (notation: \( A \sim B \)) if there exists a nonsingular matrix \( S \) such that \( B = S^{-1} A S \).
Exercise 8.2.2. Let $A$ and $B$ be similar matrices. Show that, for $k \geq 0$, we have $A^k \sim B^k$.

Proposition 8.2.3. Let $A \sim B$ (Def. 8.2.1). Then $\det A = \det B$.

Exercise 8.2.4. Let $N \in M_n(\mathbb{F})$ be a matrix. Then $N$ is nilpotent if and only if it is similar to a strictly upper triangular matrix.

Proposition 8.2.5. Every matrix is similar to a block diagonal matrix where each block has the form $\alpha I + N$ where $\alpha \in \mathbb{F}$ and $N$ is nilpotent.

Blocks of this form are “proto-Jordan blocks.” We will discuss Jordan blocks and the Jordan canonical form of matrices in Section ??

Definition 8.3.1 (Polynomial). A polynomial over the field $\mathbb{F}$ is an expression $f : \mathbb{F} \to \mathbb{F}$ of the form
\[
f = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n \quad (8.2)
\]
We may omit any terms with zero coefficient, e.g.,
\[
6 + 0t - 3t^2 + 0t^3 = 6 - 3t^2 \quad (8.3)
\]
The set of polynomials over the field $\mathbb{F}$ is denoted $\mathbb{F}[t]$.

Definition 8.3.2 (Zero polynomial). The polynomial which has all coefficients equal to zero is called the zero polynomial and is denoted by $0$.

Definition 8.3.3. The leading term of a polynomial $f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n$ is the term corresponding to the highest power of $t$ with a nonzero coefficient, that is, the term $\alpha_k t^k$ where $\alpha_k \neq 0$ and $\alpha_j = 0$ for all $j > k$. The zero polynomial does not have a leading term.

The leading coefficient of a polynomial $f$ is the coefficient of the leading term of $f$. The degree of a polynomial $f$, denoted $\deg f$, is the exponent of its leading term. A polynomial is monic if its leading coefficient is $1$.

8.3 Polynomial basics

We now digress from the theory of eigenvectors and eigenvalues in order to build a foundation in polynomials. We will discuss polynomials in more depth in Section 14.4, but now we establish the basics necessary for studying the characteristic polynomial in Section 8.4.

Definition 8.3.4 (Diagonalizability). Let $A \in M_n(\mathbb{F})$. Then $A$ is diagonalizable if $A$ is similar to a diagonal matrix.

Exercise 8.2.7.
(a) Prove that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.
(b) Prove that $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ is diagonalizable.
(c) When is the matrix $\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}$ diagonalizable?

Proposition 8.2.8. Let $A \in M_n(\mathbb{F})$. Then $A$ is diagonalizable if and only if $A$ has an eigenbasis.

\footnote{In Chapter 14 we will refine this definition and study polynomials more formally.}
CHAPTER 8. (𝔽) EIGENVECTORS AND EIGENVALUES

Convention 8.3.4. The zero polynomial has degree $-\infty$.

**Example 8.3.5.** Let $f = 6 - 3t^2 + 4t^5$. Then the leading term of $f$ is $4t^5$, the leading coefficient of $f$ is 4, and $\deg f = 5$.

**Exercise 8.3.6.** Which polynomials have degree 0?

**Notation 8.3.7.** We denote the set of polynomials of degree at most $n$ over $\mathbb{F}$ by $P_n(\mathbb{F})$.

**Definition 8.3.8 (Divisibility of polynomials).** Let $f, g \in \mathbb{F}[t]$. We say that $g$ divides $f$, or $f$ is divisible by $g$, written $g \mid f$, if there exists a polynomial $h \in \mathbb{F}[t]$ such that $f = gh$. In this case we say that $g$ is a divisor of $f$ and $f$ is a multiple of $g$.

**Definition 8.3.9 (Root of a polynomial).** Let $f \in \mathbb{F}[t]$ be a polynomial. Then $\zeta \in \mathbb{F}[t]$ is a root of $f$ if $f(\zeta) = 0$.

**Proposition 8.3.10.** Let $f \in \mathbb{F}[t]$ be a polynomial and let $\alpha \in \mathbb{F}[t]$. Then $\zeta$ is a root of $f$ if and only if $t - \zeta \mid f$.

**Definition 8.3.11 (Multiplicity of a root).** Let $f$ be a polynomial and let $\zeta$ be a root of $f$. The multiplicity of the root $\zeta$ is the largest $k$ for which $(t - \zeta)^k \mid f$.

**Proposition 8.3.12.** Let $f$ be a polynomial of degree $n$. Then $f$ has at most $n$ roots (counting multiplicity).

**Theorem 8.3.13 (Fundamental Theorem of Algebra).** Let $f \in \mathbb{C}[t]$. If $\deg f \geq 1$, then $f$ can be written as

$$f = \alpha_k \prod_{i=1}^{k} (t - \zeta_i) \quad (8.4)$$

where $\alpha_k$ is the leading coefficient of $f$ and the $\zeta_i$ are complex numbers.

8.4 The characteristic polynomial

We now establish a method for finding the eigenvalues of general $n \times n$ matrices.

**Definition 8.4.1 (Characteristic polynomial).** Let $A$ be an $n \times n$ matrix. Then the characteristic polynomial of $A$ is defined by

$$f_A(t) := \det(tI - A) \quad (8.5)$$

**Exercise 8.4.2.** (a) Explicitly calculate the characteristic polynomial for general $1 \times 1$ matrices.

(b) Explicitly calculate the characteristic polynomial for general $2 \times 2$ matrices.

**Exercise 8.4.3.** Show that the characteristic polynomial of an $n \times n$ matrix is a monic polynomial of degree $n$.

**Theorem 8.4.4.** The eigenvalues of a square matrix $A$ are precisely the roots of its characteristic polynomial.

**Proposition 8.4.5.** Every matrix $A \in M_3(\mathbb{R})$ has an eigenvector.
Exercise 8.4.6. Find a matrix $A \in M_2(\mathbb{R})$ which does not have an eigenvector.

Exercise 8.4.7. Let $N$ be an $n \times n$ nilpotent matrix. What is its characteristic polynomial?

In Section 8.1, we defined the geometric multiplicity of an eigenvalue (Def. 8.1.21). We now define the algebraic multiplicity of an eigenvalue.

Definition 8.4.8 (Algebraic multiplicity). Let $A$ be a square matrix with eigenvalue $\lambda$. The algebraic multiplicity of $\lambda$ is its multiplicity as a root of the characteristic polynomial of $A$.

Proposition 8.4.9. Let $A$ be an $n \times n$ matrix with distinct complex eigenvalues $\lambda_1, \ldots, \lambda_\ell$, and let $k_i$ be the algebraic multiplicity of $\lambda_i$ for each $i$. Then

$$\sum_{i=1}^\ell k_i = n \quad (8.6)$$

and

$$f_A = \prod_{i=1}^\ell (t - \lambda_i)^{k_i} \quad (8.7)$$

Proposition 8.4.10. Let $A$ be as in the preceding proposition. Show that the coefficient of $t^{n-\ell}$ in $f_A$ is

$$(-1)^\ell \sum_{i_1 \ldots i_\ell \text{ distinct}} \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_\ell} \quad .$$

Theorem 8.4.11. Let $A$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ (with their algebraic multiplicities). Then

$$\text{Tr } A = \sum_{i=1}^n \lambda_i \quad (8.8)$$

$$\det A = \prod_{i=1}^n \lambda_i \quad (8.9)$$

Proposition 8.4.12. Let $A$ be a square matrix and let $\lambda$ be an eigenvalue of $A$. Then the geometric multiplicity of $\lambda$ is less than or equal to the algebraic multiplicity of $\lambda$.

Exercise 8.4.13. Determine the eigenvalues and their (algebraic and geometric) multiplicities of the all ones matrix $J_n$ over $\mathbb{R}$.

Proposition 8.4.14. Let $A$ be a square matrix. Then $A$ is diagonalizable over $\mathbb{C}$ if and only if for every eigenvalue $\lambda$, the geometric and algebraic multiplicities of $\lambda$ are equal.

Proposition 8.4.15. Let $A \in M_n(\mathbb{F})$. Then the left eigenvalues of $A$ are the same as the right eigenvalues, including their geometric and algebraic multiplicities.

Proposition 8.4.16. Suppose $n \geq k$. Let $A \in \mathbb{F}^{k \times n}$ and $B \in \mathbb{F}^{n \times k}$. Then $f_{AB} = f_{BA} \cdot t^{n-k}$.

Definition 8.4.17 (Companion matrix). Let $f \in P_n[\mathbb{F}]$ be a monic polynomial, say

$$f = \alpha_0 + \alpha_1 t + \cdots + \alpha_{n-1} t^{n-1} + \alpha_n t^n \quad . (8.10)$$
Then the companion matrix of \( f \) is the matrix 
\[
C(f) \in M_n(F) \text{ defined by }
\begin{pmatrix}
0 & 0 & 0 & \cdots & 0 & -\alpha_0 \\
1 & 0 & 0 & \cdots & 0 & -\alpha_1 \\
0 & 1 & 0 & \cdots & 0 & -\alpha_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & -\alpha_{n-2} \\
0 & 0 & 0 & \cdots & 1 & -\alpha_{n-1}
\end{pmatrix}.
\tag{8.11}
\]

**Proposition 8.4.18.** Let \( f \) be a monic polynomial and let \( A = C(f) \) be its companion matrix. Then the characteristic polynomial of \( A \) is equal to \( f \). That is,
\[
f = f_A = \det(tI - C(f)).
\tag{8.12}
\]

**Corollary 8.4.19.** Every monic polynomial \( f \in \mathbb{Q}[t] \) is the characteristic polynomial of a rational matrix.

## 8.5 The Cayley-Hamilton Theorem

In Section 2.3, we defined how to substitute a square matrix into a polynomial (\( \mathfrak{I} \) Def. 2.3.3). We repeat that definition here.

**Definition 8.5.1** (Substitution of a matrix into a polynomial). Let \( f \in \mathbb{F}[t] \) be the polynomial (\( \mathfrak{I} \) Def. 8.3.1) defined by
\[
f = \alpha_0 + \alpha_1 t + \cdots + \alpha_d t^d.
\]

Just as we may substitute \( \zeta \in \mathbb{F} \) for the variable \( t \) in \( f \) to obtain a value \( f(\zeta) \in \mathbb{F} \), we may also “plug in” the matrix \( A \in M_n(\mathbb{F}) \) to obtain \( f(A) \in M_n(\mathbb{F}) \). The only thing we have to be careful about is what we do with the scalar term \( \alpha_0 \); we replace it with \( \alpha_0 \) times the identity matrix, so
\[
f(A) := \alpha_0 I + \alpha_1 A + \cdots + \alpha_d A^d.
\tag{8.13}
\]

**Exercise 8.5.2.** Let \( A \) and \( B \) be similar matrices, and let \( f \) be a polynomial. Show that \( f(A) \sim f(B) \).

The main result of this section is the following theorem.

**Theorem 8.5.3** (Cayley-Hamilton Theorem).
Let \( A \) be an \( n \times n \) matrix. Then \( f_A(A) = 0 \).

**Exercise 8.5.4.** What is wrong with the following “proof” of the Cayley-Hamilton Theorem?
\[
f_A(A) = \det(AI - A) = \det 0 = 0.
\tag{8.14}
\]

**Exercise 8.5.5.** Prove the Cayley-Hamilton Theorem by brute force
(a) for \( 1 \times 1 \) matrices,
(b) for \( 2 \times 2 \) matrices.

We will first prove the Cayley-Hamilton Theorem for diagonal matrices, and then more generally for diagonalizable matrices (\( \mathfrak{I} \) Def. 8.2.6).

**Proposition 8.5.6.** Let \( D \) be a diagonal matrix, and let \( f_D \) be its characteristic polynomial. Then \( f_D(D) = 0 \).
Proposition 8.5.7. Let $A \sim B$, so $B = S^{-1}AS$. Then

$$tI - B = S^{-1}(tI - A)S.$$  (8.15)

Theorem 8.5.8. If $A \sim B$, then the characteristic polynomials of $A$ and $B$ are equal, i.e., $f_A = f_B$. ♦

Proposition 8.5.9. Let $A$ be an $n \times n$ matrix with characteristic polynomial $f_A = \prod_{i=1}^{n}(t-\lambda_i)$. Then

$$A \sim \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$  (8.16)

Lemma 8.5.10. Let $g \in \mathbb{F}[t]$, and let $A, S \in M_n(\mathbb{F})$ with $S$ nonsingular. Then $g(S^{-1}AS) = S^{-1}g(A)S$. ♦

Corollary 8.5.11. Let $g \in \mathbb{F}[t]$ and let $A, B \in M_n(\mathbb{F})$ with $A \sim B$. Then $g(A) \sim g(B)$.

Proposition 8.5.12. The Cayley-Hamilton Theorem holds for diagonalizable matrices.

Exercise 8.5.13. Let $N$ be a nilpotent matrix. Show that $N^n = 0$.

8.6 Additional exercises

The following two exercises are easier for diagonalizable matrices.

Exercise 8.6.1. Let $f(t) = \sum_{i=0}^{\infty} \alpha_i t^i$ be a function which is convergent for all $t$ such that $|t| < r$ for some $r \in \mathbb{R}^+ \cup \{\infty\}$. Define, for $A \in M_n(\mathbb{R})$,

$$f(A) = \sum_{i=0}^{\infty} \alpha_i A^i.$$  (8.17)

Prove that $f(A)$ converges if $|\lambda_i| < r$ for all eigenvalues $\lambda_i$ of $A$. In particular, $e^A$ always converges.

Exercise 8.6.2.

(a) Find square matrices $A$ and $B$ such that $e^{A+B} \neq e^A e^B$.

(b) Give a natural sufficient condition under which $e^{A+B} = e^A e^B$. 


Chapter 9

(ℝ) Orthogonal Matrices

9.1 Orthogonal matrices

In Section 1.4, we defined the standard dot product (Def. 1.4.1) and the notions of orthogonal vectors (Def. 1.4.8). In this chapter we study orthogonal matrices, matrices whose columns form an orthonormal basis (Def. 1.5.6) of ℝⁿ.

Definition 9.1.1 (Orthogonal matrix). The matrix A ∈ Mⁿ(ℝ) is orthogonal if AᵀA = I. The set of orthogonal n × n matrices is denoted by O(n).

Fact 9.1.2. A ∈ Mⁿ(ℝ) is orthogonal if and only if its columns form an orthonormal basis of ℝⁿ.

Proposition 9.1.3. O(n) is a group (Def. 14.2.1) under matrix multiplication (it is called the orthogonal group).

Exercise 9.1.4. Which diagonal matrices are orthogonal?

Theorem 9.1.5 (Third Miracle of Linear Algebra). Let A ∈ Mⁿ(ℝ). Then the columns of A are orthonormal if and only if the rows of A are orthonormal. ♦

Proposition 9.1.6. Let A ∈ O(n). Then all eigenvalues of A have absolute value 1.

Exercise 9.1.7. The matrix A ∈ Mⁿ(ℝ) is orthogonal if and only if A preserves the dot product, i.e., for all v, w ∈ ℝⁿ, we have (Av)ᵀ(Aw) = vᵀw.

Exercise 9.1.8. The matrix A ∈ Mⁿ(ℝ) is orthogonal if and only if A preserves the norm, i.e., for all v ∈ ℝⁿ, we have ||Av|| = ||v||.

9.2 Orthogonal similarity

Definition 9.2.1 (Orthogonal similarity). Let A, B ∈ Mⁿ(ℝ). We say that A is orthogonally similar to B, denoted A ∼ o B, if there exists an orthogonal matrix O such that A = O⁻¹BO.

Note that O⁻¹BO = OᵀBO because O is orthogonal.

Proposition 9.2.2. Let A ∼ o B.

(a) If A is symmetric then so is B.

(b) If A is orthogonal then so is B.

Proposition 9.2.3. Let A ∼ o diag(λ₁, ..., λₙ). Then
9.3. ADDITIONAL EXERCISES

(a) If all eigenvalues of \( A \) are real then \( A \) is symmetric.

(b) If all eigenvalues of \( A \) have unit absolute value then \( A \) is orthogonal.

**Proposition 9.2.4.** \( A \in M_n(\mathbb{R}) \) has an orthonormal eigenbasis if and only if \( A \) is orthogonally similar to a diagonal matrix.

**Proposition 9.2.5.** Let \( A \in M_n(\mathbb{R}) \). Then \( A \in O(n) \) if and only if it is orthogonally similar to a matrix which is the diagonal sum of some of the following: an identity matrix, a negative identity matrix, \( 2 \times 2 \) rotation matrices (compare with Prop. [16.4.45]).

**Examples 9.2.6.** The following are examples of the matrices described in the preceding proposition.

\[
\begin{pmatrix}
-1 & 0 & 0 \\
0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}
\end{pmatrix}
\]

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2}
\end{pmatrix}
\]

9.3 Additional exercises

**Definition 9.3.1 (Hadamard matrix).** The matrix \( A = (\alpha_{ij}) \in M_n(\mathbb{R}) \) is an Hadamard matrix if \( \alpha_{ij} = \pm 1 \) for all \( i, j \), and the columns of \( A \) are orthogonal. We denote by \( \mathcal{H} \) the set

\[\mathcal{H} := \{ n \mid \text{an } n \times n \text{ Hadamard matrix exists} \} . \]

(9.1)

**Example 9.3.2.** The matrix \( \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \) is an Hadamard matrix.

**Exercise 9.3.3.** Let \( A \) be an \( n \times n \) Hadamard matrix. Create a \( 2n \times 2n \) Hadamard matrix.

**Proposition 9.3.4.** If \( n \in \mathcal{H} \) and \( n > 2 \), then \( 4 \mid n \).

**Proposition 9.3.5.** If \( p \) is prime and \( p \equiv -1 \pmod{4} \), then \( p + 1 \in \mathcal{H} \).

**Proposition 9.3.6.** If \( k, \ell \in \mathcal{H} \), then \( k \ell \in \mathcal{H} \).

**Proposition 9.3.7.** Let \( A \) be an \( n \times n \) Hadamard matrix. Then \( \frac{1}{\sqrt{n}} A \) is an orthogonal matrix.
Chapter 10

(R) The Spectral Theorem

10.1 Statement of the Spectral Theorem

The Spectral Theorem is one of the most significant results of linear algebra, as well as one of the most frequently applied mathematical results in pure math, applied math, and science. We will see a number of different versions of this theorem, and we will not prove it until Section 19.4. However, we now have developed the tools necessary to understand the statement of the theorem and some of its applications.

Theorem 10.1.1 (The Spectral Theorem for real symmetric matrices). Let $A \in M_n(\mathbb{R})$ be a real symmetric matrix. Then $A$ has an orthonormal eigenbasis.

The Spectral Theorem can be restated in terms of orthogonal similarity (Def. 9.2.1).

Theorem 10.1.2 (The Spectral Theorem for real symmetric matrices, restated). Let $A \in M_n(\mathbb{R})$ be a real symmetric matrix. Then $A$ is orthogonally similar to a diagonal matrix.

Exercise 10.1.3. Verify that these two formulations of the Spectral Theorem are equivalent.

Corollary 10.1.4. Let $A$ be a real symmetric matrix. Then $A$ is diagonalizable.

Corollary 10.1.5. Let $A$ be a real symmetric matrix. Then all of the eigenvalues of $A$ are real.

10.2 Applications of the Spectral Theorem

Although we have not yet proved the Spectral Theorem, we can already begin to study some of its many applications.

Proposition 10.2.1. If two symmetric matrices are similar then they are orthogonally similar.

Exercise 10.2.2. Let $A$ be a symmetric real $n \times n$ matrix, and let $v \in \mathbb{R}^n$. Let $b = (b_1, \ldots, b_n)$ be an orthonormal eigenbasis of $A$. Express $v^T A v$ in terms of the eigenvalues and the coordinates of $v$ with respect to $b$.

Definition 10.2.3 (Positive definite matrix). An $n \times n$ real matrix $A \in M_n(\mathbb{R})$ is positive definite if for all $x \in \mathbb{R}^n$ ($x \neq 0$), we have $x^T A x > 0$. 

Proposition 10.2.4. Let $A \in M_n(\mathbb{R})$ be a real symmetric $n \times n$ matrix. Then $A$ is positive definite if and only if all eigenvalues of $A$ are positive.

Another consequence of the Spectral Theorem is Rayleigh’s Principle.

Definition 10.2.5 (Rayleigh quotient). Let $A \in M_n(\mathbb{R})$. The Rayleigh quotient of $A$ is a function $R_A : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ defined by

$$R_A(v) = \frac{v^T A v}{\|v\|^2}. \quad (10.1)$$

Recall that $\|v\|^2 = v^T v$ (Def. 1.5.2).

Proposition 10.2.6 (Rayleigh’s Principle). Let $A$ be an $n \times n$ real symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then

(a) $\max_{v \in \mathbb{R}^n \setminus \{0\}} R_A(v) = \lambda_1$

(b) $\min_{v \in \mathbb{R}^n \setminus \{0\}} R_A(v) = \lambda_n$

Theorem 10.2.7 (Courant-Fischer). Let $A$ be a symmetric real matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then

$$\lambda_i = \max_{U \subseteq \mathbb{R}^n} \min_{v \in U \setminus \{0\}} R_A(v). \quad (10.2)$$

Theorem 10.2.8 (Interlacing). Let $A \in M_n(\mathbb{R})$ be an $n \times n$ symmetric real matrix. Let $B$ be the $(n-1) \times (n-1)$ matrix obtained by deleting the $i$-th column and the $i$-th row of $A$ (so $B$ is also symmetric). Prove that the eigenvalues of $A$ and $B$ interlace, i.e., if the eigenvalues of $A$ are $\lambda_1 \geq \cdots \geq \lambda_n$ and the eigenvalues of $B$ are $\mu_1 \geq \cdots \geq \mu_{n-1}$, then

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n.$$

$\diamondsuit$
Chapter 11

(\mathbb{F}, \mathbb{R}) Bilinear and Quadratic Forms

11.1 (\mathbb{F}) Linear and bilinear forms

Definition 11.1.1 (Linear form). A linear form is a function \( f : \mathbb{F}^n \rightarrow \mathbb{F} \) with the following properties.

(a) \( f(x + y) = f(x) + f(y) \) for all \( x, y \in \mathbb{F}^n \);

(b) \( f(\lambda x) = \lambda f(x) \) for all \( x \in \mathbb{F}^n \) and \( \lambda \in \mathbb{F} \).

Exercise 11.1.2. Let \( f \) be a linear form. Show that \( f(0) = 0 \).

Definition 11.1.3 (Dual space). The set of linear forms \( f : \mathbb{F}^n \rightarrow \mathbb{F} \) is called the dual space of \( \mathbb{F}^n \) and is denoted \( (\mathbb{F}^n)^* \).

Example 11.1.4. The function \( f(x) = x_1 + \cdots + x_n \) (where \( x = (x_1, \ldots, x_n)^T \)) is a linear form. More generally, for any \( a = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{F}^n \), the function

\[
f(x) = a^T x = \sum_{i=1}^{n} \alpha_i x_i \tag{11.1}
\]

is a linear form.

Theorem 11.1.5 (Representation Theorem for Linear Forms). Every linear form \( f : \mathbb{F}^n \rightarrow \mathbb{F} \) has the form (11.1) for some column vector \( a \in \mathbb{F}^n \).

Definition 11.1.6 (Bilinear form). A bilinear form is a function \( f : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{F} \) with the following properties.

(a) \( f(x_1 + x_2, y) = f(x_1, y) + f(x_2, y) \)

(b) \( f(\lambda x, y) = \lambda f(x, y) \)

(c) \( f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2) \)

(d) \( f(x, \lambda y) = \lambda f(x, y) \)

Exercise 11.1.7. Let \( f : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{F} \) be a bilinear form. Show that for all \( x, y \in \mathbb{F} \), we have

\[
f(x, 0) = f(0, y) = 0 \tag{11.2}
\]

The next result explains the term “bilinear.”

Proposition 11.1.8. The function \( f : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{F} \) is a bilinear form exactly if

(a) for all \( a \in \mathbb{F}^n \) the function \( f_a^{(1)} : \mathbb{F}^n \rightarrow \mathbb{F} \) defined by \( f_a^{(1)}(x) = f(a, x) \) is a linear form and

(b) for all \( b \in \mathbb{F}^n \) the function \( f_b^{(2)} : \mathbb{F}^n \rightarrow \mathbb{F} \) defined by \( f_b^{(2)}(x) = f(x, b) \) is a linear form.
11.2  \((\mathbb{F})\) Multivariate Polynomials

**Examples 11.1.9.** The standard dot product

\[ f(x, y) = x^T y = \sum_{i=1}^{n} x_i y_i \quad (11.3) \]

is a bilinear form. More generally, for any matrix \(A \in M_n(\mathbb{F})\), the expression

\[ f(x, y) = x^T A y = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} x_i y_j \quad (11.4) \]

is a bilinear form.

**Theorem 11.1.10** (Representation Theorem for bilinear forms). Every bilinear form \(f\) has the form \((11.4)\) for some matrix \(A\). \[\Box\]

**Definition 11.1.11** (Nonsingular bilinear form). We say that the bilinear form \(f(x, y) = x^T A y\) is nonsingular if the matrix \(A\) is nonsingular.

**Exercise 11.1.12.** Is the standard dot product nonsingular?

**Exercise 11.1.13.**

(a) Assume \(\mathbb{F}\) does not have characteristic 2, i.e., \(1 + 1 \neq 0\) in \(\mathbb{F}\) (\(|\equiv \) Section 14.3 for more about the characteristic of a field). Prove: if \(n\) is odd and the bilinear form \(f : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}\) satisfies \(f(x, x) = 0\) for all \(x \in \mathbb{F}^n\), then \(f\) is singular.

(b)* Prove this without the assumption that \(1 + 1 \neq 0\).

(b) Over every field \(\mathbb{F}\), find a nonsingular bilinear form \(f\) over \(\mathbb{F}^2\) such that \(f(x, x) = 0\) for all \(x \in \mathbb{F}^2\).

(c) Extend this to all even dimensions.

**11.2  \((\mathbb{F})\) Multivariate Polynomials**

**Definition 11.2.1** (Multivariate monomial). A monomial in the variables \(x_1, \ldots, x_n\) is an expression of the form

\[ f = \alpha \prod_{i=1}^{n} x_i^{k_i} \]

for some nonzero scalar \(\alpha\) and exponents \(k_i\).

The degree of this monomial is

\[ \deg f := \sum_{i=1}^{n} k_i \quad (11.5) \]

**Examples 11.2.2.** The following expressions are multivariate monomials of degree 4 in the variables \(x_1, \ldots, x_6\): \(x_1^2 x_2 x_3, 5x_4 x_5^3, -3x_6^4\).

**Definition 11.2.3** (Multivariate polynomial). A multivariate polynomial is an expression which is the sum of multivariate monomials. Observe that the preceding definition includes the possibility of the empty sum, corresponding to the 0 polynomial.

**Definition 11.2.4** (Monic monomial). We call the monomials of the form \(\prod_{i=1}^{n} x_i^{k_i}\) monic. We define the monic part of the monomial \(f = \alpha \prod_{i=1}^{n} x_i^{k_i}\) to be the monomial \(\prod_{i=1}^{n} x_i^{k_i}\).
Definition 11.2.5 (Standard form of a multivariate polynomial). A multivariate polynomial is in standard form if it is expressed as a (possibly empty) sum of monomials with distinct monic parts. The empty sum of monomials is the zero polynomial, denoted by 0.

Examples 11.2.6. The following are multivariate polynomials in the variables $x_1, \ldots, x_7$ (in standard form).

(a) $3x_1^3x_3^3 + 2x_2x_6 - x_4$
(b) $4x_1^3 + 2x_5x_7^2 + 3x_1x_5x_7$
(c) 0

Definition 11.2.7 (Degree of a multivariate polynomial). The degree of a multivariate polynomial $f$ is the highest degree of a monomial in the standard form expression for $f$. The degree of the 0 polynomial is defined to be $-\infty$.

Exercise 11.2.8. Let $f$ and $g$ be multivariate polynomials. Then

(a) $\deg(f + g) \leq \max\{\deg f, \deg g\}$
(b) $\deg(fg) = \deg f + \deg g$

Note that by our convention, these rules remain valid if $f$ or $g$ is the zero polynomial.

Definition 11.2.9 (Homogeneous multivariate polynomial). The multivariate polynomial $f$ is a homogeneous polynomial of degree $k$ if every monomial in the standard form expression of $f$ has degree $k$.

Fact 11.2.10. The 0 polynomial is a homogeneous polynomial of degree $k$ for all $k$.

Fact 11.2.11.

(a) If $f$ and $g$ are homogeneous polynomials of degree $k$, then $f + g$ is a homogeneous polynomial of degree $k$.

(b) If $f$ is a homogeneous polynomial of degree $k$ and $g$ is a homogeneous polynomial of degree $\ell$, then $fg$ is a homogeneous polynomial of degree $k + \ell$.

Exercise 11.2.12. For all $n$ and $k$, count the monic monomials of degree $k$ in the variables $x_1, \ldots, x_n$. Note that this is the dimension (Def. 15.3.6) of the space of homogeneous polynomials of degree $k$.

Exercise 11.2.13. What are the homogeneous polynomials of degree 0?

Fact 11.2.14. Linear forms are exactly the homogeneous polynomials of degree 1.

In the next section we explore quadratic forms, which are the homogeneous polynomials of degree 2 (Def. 11.6).
11.3 \((\mathbb{R})\) Quadratic forms

In this section we restrict our attention to the field of real numbers.

**Definition 11.3.1 (Quadratic form).** A quadratic form is a function \(Q : \mathbb{R}^n \rightarrow \mathbb{R}^n\) where \(Q(x) = f(x,x)\) for some bilinear form \(f\).

**Definition 11.3.2.** Let \(A \in M_n(\mathbb{R})\) be an \(n \times n\) matrix. The quadratic form associated with \(A\), denoted \(Q_A\), is defined by
\[
Q_A(x) = x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j .
\] (11.6)

Note that for all matrices \(A \in M_n(\mathbb{R})\), we have \(Q_A(0) = 0\). Note further that for a symmetric matrix \(B\), the quadratic form \(Q_B\) is the numerator of the Rayleigh quotient \(R_B\) (Def. 10.2.5).

**Proposition 11.3.3.** For all \(A \in M_n(\mathbb{R})\), there is a unique \(B \in M_n(\mathbb{R})\) such that \(B\) is a symmetric matrix and \(Q_A = Q_B\), i.e., for all \(x \in \mathbb{R}^n\), we have
\[
x^T A x = x^T B x .
\]

**Exercise 11.3.4.** Find the symmetric matrix \(B\) such that
\[
Q_B(x) = 3x_1^2 - 7x_1x_2 + 2x_2^2
\]
where \(x = (x_1, x_2)^T\).

**Definition 11.3.5.** Let \(Q\) be a quadratic form in \(n\) variables.

(a) \(Q\) is **positive definite** if \(Q(x) > 0\) for all \(x \neq 0\).

(b) \(Q\) is **positive semidefinite** if \(Q(x) \geq 0\) for all \(x\).

(c) \(Q\) is **negative definite** if for all \(Q(x) < 0\) for all \(x \neq 0\).

(d) \(Q\) is **negative semidefinite** if for all \(Q(x) \leq 0\) for all \(x\).

(e) \(Q\) is **indefinite** if it is neither positive semidefinite nor negative semidefinite, i.e., there exist \(x, y \in \mathbb{R}^n\) such that \(Q(x) > 0\) and \(Q(y) < 0\).

**Definition 11.3.6.** We say that a matrix \(A \in M_n(\mathbb{R})\) is **positive definite** if it is symmetric and its associated quadratic form \(Q_A\) is positive definite. Positive semidefinite, negative definite, negative semidefinite, and indefinite symmetric matrices are defined analogously.

Notice that we shall not call a non-symmetric matrix \(A\) positive definite, etc., even if the quadratic form \(Q_A\) is positive definite, etc.

**Exercise 11.3.7.** Categorize diagonal matrices according to their definiteness, i.e., tell, which diagonal matrices are positive definite, etc.

**Exercise 11.3.8.** Let \(A \in M_n(\mathbb{R})\). Assume \(Q_A\) is positive definite and \(\lambda\) is a real eigenvalue of \(A\). Prove \(\lambda > 0\). Note that \(A\) is not necessarily symmetric.
Corollary 11.3.9. If $A$ is a positive definite (and therefore symmetric by definition) matrix and $\lambda$ is an eigenvalue of $A$ then $\lambda > 0$.

Proposition 11.3.10. Let $A \in M_n(\mathbb{R})$ be a symmetric real matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$.

(a) $Q_A$ is positive definite if and only if $\lambda_i > 0$ for all $i$.

(b) $Q_A$ is positive semidefinite if and only if $\lambda_i \geq 0$ for all $i$.

(c) $Q_A$ is negative definite if and only if $\lambda_i < 0$ for all $i$.

(d) $Q_A$ is negative semidefinite if and only if $\lambda_i \leq 0$ for all $i$.

(e) $Q_A$ is indefinite if there exist $i$ and $j$ such that $\lambda_i > 0$ and $\lambda_j < 0$ for all $i$.

Proposition 11.3.11. If $A \in M_n(\mathbb{R})$ is positive definite, then its determinant is positive.

Exercise 11.3.12. Show that if $A$ and $B$ are symmetric $n \times n$ matrices and $A \sim B$ and $A$ is positive definite, then so is $B$.

Definition 11.3.13 (Corner matrix). Let $A = (\alpha_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$ and define for $k = 1, \ldots, n$, the corner matrix $A_k := (\alpha_{ij})_{i,j=1}^k$ to be the $k \times k$ submatrix of $A$, obtained by taking the intersection of the first $k$ rows of $A$ with the first $k$ columns of $A$. In particular, $A_n = A$. The $k$-th corner determinant of $A$ is $\det A_k$.

Exercise 11.3.14. Let $A \in M_n(\mathbb{R})$ be a symmetric matrix.

(a) If $A$ is positive definite then all of its corner matrices are positive definite.

(b) If $A$ is positive definite then all of its corner determinants are positive.

The next theorem says that (b) is actually a necessary and sufficient condition for positive definiteness.

Theorem 11.3.15. Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. $A$ is positive definite if and only if all of its corner determinants are positive.

Exercise 11.3.16. Show that the following statement is false for every $n \geq 2$: The $n \times n$ symmetric matrix $A$ is positive semidefinite if and only if all corner determinants are nonnegative.

Exercise 11.3.17. Show that the statement of Ex. 11.3.16 remains false for $n \geq 3$ if in addition we require all diagonal elements to be positive.

Exercise 11.3.18. Let $A \in M_n(\mathbb{R})$ be an $n \times n$ matrix such that $Q_A$ is positive definite. Show that $A$ need not have all of its corner determinants positive. Give a $2 \times 2$ counterexample. Contrast this with part (b) of Ex. 11.3.14.
Proposition 11.3.19. Let \( A \in M_n(\mathbb{R}) \). If \( Q_A \) is positive definite, then so is \( Q_{A^T} \); analogous results hold for positive semidefinite, negative definite, negative semidefinite, and indefinite quadratic forms.

Exercise 11.3.20. Let \( A, B, C \in M_n(\mathbb{R}) \). Assume \( B = C^T A C \). Prove: if \( Q_A \) is positive semidefinite then so is \( Q_B \).

Exercise 11.3.21. Let \( A, B \in M_n(\mathbb{R}) \). Show that if \( A \sim_o B \) (\( A \) and \( B \) are orthogonally similar, Def. 9.2.1) and \( Q_A \) is positive definite then \( Q_B \) is positive definite.

Exercise 11.3.22. Let \( A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \). Prove that \( Q_A \) is positive definite if and only if \( \alpha, \delta > 0 \) and \( (\beta + \gamma)^2 < 4 \alpha \delta \).

Exercise 11.3.23. Let \( A, B \in M_n(\mathbb{R}) \) be matrices that are not necessarily symmetric.

(a) Prove that if \( A \sim B \) and \( Q_A \) is positive definite, then \( Q_B \) cannot be negative definite.

(b) Find \( 2 \times 2 \) matrices \( A \) and \( B \) such that \( A \sim B \) and \( Q_A \) is positive definite but \( Q_B \) is indefinite. (Contrast this with Ex. 11.3.12)

11.4 (\( \mathbb{F} \)) Geometric algebra (optional)

Let us fix a bilinear form \( f : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{F} \).

Definition 11.4.1 (Orthogonality). Let \( x, y \in \mathbb{F}^n \). We say that \( x \) and \( y \) are orthogonal with respect to \( f \) (notation: \( x \perp y \)) if \( f(x, y) = 0 \).

Definition 11.4.2. Let \( S, T \subseteq \mathbb{F}^n \). For \( v \in \mathbb{F}^n \), we say that \( v \) is orthogonal to \( S \) (notation: \( v \perp S \)) if for all \( s \in S \), we have \( v \perp s \). Moreover, we say that \( S \) is orthogonal to \( T \) (notation: \( S \perp T \)) if \( s \perp t \) for all \( s \in S \) and \( t \in T \).

Definition 11.4.3. Let \( S \subseteq \mathbb{F}^n \). Then \( S^\perp \) ("\( S \) perp") is the set of vectors orthogonal to \( S \), i.e.,

\[
S^\perp := \{ v \in \mathbb{F}^n \mid v \perp S \}.
\] (11.7)

Proposition 11.4.4. For all subsets \( S \subseteq \mathbb{F}^n \), we have \( S^\perp \subseteq \mathbb{F}^n \).

Proposition 11.4.5. Let \( S \subseteq \mathbb{F}^n \). Then \( S \subseteq (S^\perp)^\perp \).

Exercise 11.4.6. Prove: \( (\mathbb{F}^n)^\perp = \{0\} \) if and only if \( f \) is nonsingular.

Exercise 11.4.7. Verify \( \{0\}^\perp = \mathbb{F}^n \).

Exercise 11.4.8. What is \( \emptyset^\perp \)?

\[
\ast \ast \ast
\]

For the rest of this section we assume that \( f \) is nonsingular.

Theorem 11.4.9 (Dimensional complementarity). Let \( U \subseteq \mathbb{F}^n \). Then

\[
\dim U + \dim U^\perp = n .
\] (11.8)

\[\diamondsuit\]
Corollary 11.4.10. Let $S \subseteq \mathbb{F}^n$. Then
\[(S^\perp)^\perp = \text{span}(S) . \]  
(11.9)
In particular, if $U \subseteq \mathbb{F}^n$ then
\[(U^\perp)^\perp = U . \]  
(11.10)

Definition 11.4.11 (Isotropic vector). The vector $v \in \mathbb{F}^n$ is isotropic if $v \neq 0$ and $v \perp v$.

Exercise 11.4.12. Let $A = I$, i.e., $f(x, y) = x^T y$ (the standard dot product).
(a) Prove: over $\mathbb{R}$, there are no isotropic vectors.
(b) Find isotropic vectors in $\mathbb{C}^2$, $\mathbb{F}_5^2$, and $\mathbb{F}_2^2$.
(c) For what primes $p$ is there an isotropic vector in $\mathbb{F}_p^2$?

Definition 11.4.13 (Totally isotropic subspace). The subspace $U \subseteq \mathbb{F}^n$ is totally isotropic if $U \perp U$, i.e., $U \subseteq U^\perp$.

Exercise 11.4.14. If $S \subseteq \mathbb{F}^n$ and $S \perp S$ then $\text{span}(S)$ is a totally isotropic subspace.

Corollary 11.4.15 (to Theorem 11.4.9). If $U \subseteq \mathbb{F}^n$ is a totally isotropic subspace then $\dim U \leq \lfloor \frac{n}{2} \rfloor$.

Exercise 11.4.16. For even $n$, find an $\frac{n}{2}$-dimensional totally isotropic subspace in $\mathbb{C}^n$, $\mathbb{F}_5^n$, and $\mathbb{F}_2^n$.

Exercise 11.4.17. Let $\mathbb{F}$ be a field and let $k \geq 2$. Consider the following statement.

\[\text{Stm}(\mathbb{F}, k): \text{If } U \subseteq \mathbb{F}^n \text{ is a } k\text{-dimensional subspace then } U \text{ contains an isotropic vector.}\]

Prove:
(a) if $k \leq \ell$ and Stm($\mathbb{F}, k$) is true, then Stm($\mathbb{F}, \ell$) is also true;
(b) Stm($\mathbb{F}, 2$) is true
   (b1) for $\mathbb{F} = \mathbb{F}_2$ and
   (b2) for $\mathbb{F} = \mathbb{C}$;
(c) Stm($\mathbb{F}, 2$) holds for all finite fields of characteristic 2;
(d) Stm($\mathbb{F}, 2$) is false for all finite fields of odd characteristic;
(e)* Stm($\mathbb{F}, 3$) holds for all finite fields.

Exercise 11.4.18. Prove: if Stm($\mathbb{F}, 2$) is true then every maximal totally isotropic subspace of $\mathbb{F}^n$ has dimension $\lfloor \frac{n}{2} \rfloor$. In particular, this conclusion holds for $\mathbb{F} = \mathbb{F}_2$ and for $\mathbb{F} = \mathbb{C}$.

Exercise 11.4.19. Prove: for all finite fields $\mathbb{F}$, every maximal totally isotropic subspace of $\mathbb{F}^n$ has dimension $\geq \frac{n}{2} - 1$. 
Chapter 12

(C) Complex Matrices

12.1 Complex numbers

Before beginning our discussion of matrices with entries taken from the field $\mathbb{C}$, we provide a refresher of complex numbers and their properties.

Definition 12.1.1 (Complex number). A complex number is a number $z \in \mathbb{C}$ of the form $z = a + bi$, where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$.

Notation 12.1.2. Let $z$ be a complex number. Then $\text{Re } z$ denotes the real part of $z$ and $\text{Im } z$ denotes the imaginary part of $z$. In particular, if $z = a + bi$, then $\text{Re } z = a$ and $\text{Im } z = b$.

Definition 12.1.3 (Complex conjugate). Let $z = a + bi \in \mathbb{C}$. Then the complex conjugate of $z$, denoted $\overline{z}$, is

$$\overline{z} = a - bi.$$  

(12.1)

Exercise 12.1.4. Let $z_1, z_2 \in \mathbb{C}$. Show that

$$\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}.$$  

(12.2)

Fact 12.1.5. Let $z = a + bi$. Then $z \overline{z} = a^2 + b^2$. In particular, $z \overline{z} \in \mathbb{R}$ and $z \overline{z} \geq 0$.

Definition 12.1.6 (Magnitude of a complex number). Let $z \in \mathbb{C}$. Then the magnitude, norm, or absolute value of $z$ is $|z| = \sqrt{z \overline{z}}$. If $|z| = 1$, then $z$ is said to have unit norm.

Proposition 12.1.7. Let $z \in \mathbb{C}$ have unit norm. Then $z$ can be expressed in the form

$$z = \cos \theta + i \sin \theta$$  

(12.3)

for some $\theta \in [0, 2\pi)$.

Proposition 12.1.8. The complex number $z$ has unit norm if and only if $\overline{z} = z^{-1}$.

Until now, we have dealt with matrices over the reals or over a general field $\mathbb{F}$. We now turn our attention specifically to matrices with complex entries.

12.2 Hermitian dot product in $\mathbb{C}^n$

Definition 12.2.1 (Conjugate-transpose). Let $A = (a_{ij}) \in \mathbb{C}^{k \times \ell}$. The conjugate-transpose of $A$ is the $\ell \times k$ matrix $A^*$ whose $(i, j)$ entry is $\overline{a_{ji}}$. The conjugate-transpose is also called the (Hermitian) adjoint.
CHAPTER 12. (C) COMPLEX MATRICES

Example 12.2.2. The conjugate-transpose of the matrix

\[
\begin{pmatrix}
2 & 3 - 4i & -2i \\
1 + i & -5 & 6i \\
\end{pmatrix}
\]

is the matrix

\[
\begin{pmatrix}
2 & 1 - i \\
3 + 4i & -5 \\
2i & -6i \\
\end{pmatrix}
\]

Fact 12.2.3. Let \( A, B \in \mathbb{C}^{k \times n} \). Then

\[(A + B)^* = A^* + B^* . \tag{12.4}\]

Exercise 12.2.4. Let \( A \in \mathbb{C}^{k \times n} \) and let \( B \in \mathbb{C}^{n \times \ell} \). Show that \((AB)^* = B^* A^* \).

Fact 12.2.5. Let \( \lambda \in \mathbb{C} \) and \( A \in \mathbb{C}^{k \times n} \). Then

\[(\lambda A)^* = \bar{\lambda} A^* . \]

In Section 1.4, we defined the standard dot product (\( \mathbb{R} \) Def. 1.4.1). We now define the standard Hermitian dot product for vectors in \( \mathbb{C}^n \).

Definition 12.2.6 (Standard Hermitian dot product). Let \( v, w \in \mathbb{C}^n \). Then the Hermitian dot product of \( v \) with \( w \) is

\[v \cdot w := v^* w = \sum_{i=1}^{n} \bar{\alpha}_i \beta_i \tag{12.5}\]

where \( v = (\alpha_1, \ldots, \alpha_n)^T \) and \( w = (\beta_1, \ldots, \beta_n)^T \).

In particular, observe that \( v^* v \) is real and positive for all \( v \neq 0 \). The following pair of exercises show some of the things that would go wrong if we did not conjugate.

Exercise 12.2.7. Find a nonzero vector \( v \in \mathbb{C}^n \) such that \( v^T v = 0 \).

Exercise 12.2.8. Let \( A \in \mathbb{C}^{k \times n} \).

(a) Show that \( \text{rk} (A^* A) = \text{rk} A \).

(b) Find \( A \in M_2(\mathbb{C}) \) such that \( \text{rk} (A^T A) < \text{rk} (A) \).

The standard dot product in \( \mathbb{R}^n \) (\( \mathbb{R} \) Def. 1.4.1) carries with it the notions of norm and orthogonality; likewise, the standard Hermitian dot product carries these notions with it.

Definition 12.2.9 (Norm). Let \( v \in \mathbb{C}^n \). The norm of \( v \), denoted \( \|v\| \), is

\[\|v\| := \sqrt{v \cdot v} . \tag{12.6}\]

This norm is also referred to as the (complex) Euclidean norm or the \( \ell^2 \) norm.

Fact 12.2.10. If \( v = (\alpha_1, \ldots, \alpha_n)^T \), then

\[\|v\| = \sqrt{\sum_{i=1}^{n} |\alpha_i|^2} . \tag{12.7}\]

Definition 12.2.11 (Orthogonality). The vectors \( v, w \in \mathbb{C}^n \) are orthogonal (notation: \( x \perp y \)) if \( v \cdot w = 0 \).

Definition 12.2.12 (Orthogonal system). An orthogonal system in \( \mathbb{C}^n \) is a list of (pairwise) orthogonal nonzero vectors in \( \mathbb{C}^n \).

Exercise 12.2.13. Let \( S \subseteq \mathbb{C}^k \) be an orthogonal system in \( \mathbb{C}^n \). Prove that \( S \) is linearly independent.
12.3. HERMITIAN AND UNITARY MATRICES

Definition 12.2.14 (Orthonormal system). An orthonormal system in \( \mathbb{C}^n \) is a list of (pairwise) orthogonal vectors in \( V \), all of which have unit norm. So \( (v_1, v_2, \ldots) \) is an orthonormal system if \( v_i \cdot v_j = \delta_{ij} \) for all \( i, j \).

Definition 12.2.15 (Orthonormal basis). An orthonormal basis of \( \mathbb{C}^n \) is an orthonormal system that is a basis of \( \mathbb{C}^n \).

12.3 Hermitian and unitary matrices

Definition 12.3.1 (Self-adjoint matrix). The matrix \( A \in M_n(\mathbb{C}) \) is said to be self-adjoint or Hermitian if \( A^* = A \).

Example 12.3.2. The matrix \( \begin{pmatrix} 8 & 2 + 5i \\ 2 - 5i & -3 \end{pmatrix} \) is self-adjoint.

Exercise 12.3.3. Which diagonal matrices are Hermitian?

Theorem 12.3.4. Let \( A \) be a Hermitian matrix. Then all eigenvalues of \( A \) are real.

Exercise 12.3.5 (Alternative proof of the real Spectral Theorem). The key part of the proof of the Spectral Theorem (\[
\text{Theorem 19.4.4}\]
given in Section 19.4) is the following lemma (\[
\text{Lemma 19.4.7}\]): Let \( A \) be a symmetric real matrix. Then \( A \) has an eigenvector. Derive this lemma from Theorem 12.3.4

Recall that a square matrix \( A \) is orthogonal (\[
\text{Def. 9.1.1}\]) if \( A^T A = I \). Unitary matrices are the complex generalization of orthogonal matrices.

Definition 12.3.6 (Unitary matrix). The matrix \( A \in M_n(\mathbb{C}) \) is unitary if \( A^* A = I \). The set of unitary \( n \times n \) matrices is denoted by \( U(n) \).

Fact 12.3.7. \( A \in M_n(\mathbb{C}) \) is unitary if and only if its columns form an orthonormal basis of \( \mathbb{C}^n \).

Proposition 12.3.8. \( U(n) \) is a group under matrix multiplication (it is called the unitary group).

Exercise 12.3.9. Which diagonal matrices are unitary?

Theorem 12.3.10 (Third Miracle of Linear Algebra). Let \( A \in M_n(\mathbb{C}) \). Then the columns of \( A \) are orthonormal if and only if the rows of \( A \) are orthonormal.

Proposition 12.3.11. Let \( A \in U(n) \). Then all eigenvalues of \( A \) have absolute value 1.

Exercise 12.3.12. The matrix \( A \in M_n(\mathbb{C}) \) is unitary if and only if \( A \) preserves the Hermitian dot product, i.e., for all \( v, w \in \mathbb{C}^n \), we have \( (Av)^*(Aw) = v^*w \).

Exercise* 12.3.13. The matrix \( A \in M_n(\mathbb{C}) \) is unitary if and only if \( A \) preserves the norm, i.e., for all \( v \in \mathbb{C}^n \), we have \( \|Av\| = \|v\| \).

Warning. The proof of this is trickier than in the real case (\[
\text{Ex. 9.1.8}\]).
12.4 Normal matrices and unitary similarity

Definition 12.4.1 (Normal matrix). A matrix \( A \in M_n(\mathbb{C}) \) is normal if it commutes with its conjugate-transpose, i.e., \( AA^* = A^*A \).

Exercise 12.4.2. Which diagonal matrices are normal?

Definition 12.4.3 (Unitary similarity). Let \( A, B \in M_n(\mathbb{C}) \). We say that \( A \) is unitarily similar to \( B \), denoted \( A \sim_u B \), if there exists a unitary matrix \( U \) such that \( A = U^{-1}BU \).

Note that \( U^{-1}BU = U^*BU \) because \( U \) is unitary.

Proposition 12.4.4. Let \( A \sim_u B \).

(a) If \( A \) is Hermitian then so is \( B \).

(b) If \( A \) is unitary then so is \( B \).

(c) If \( A \) is normal then so is \( B \).

Proposition 12.4.5. Let \( A \sim_u \text{diag}(\lambda_1, \ldots, \lambda_n) \). Then

(a) \( A \) is normal;

(b) if all eigenvalues of \( A \) are real then \( A \) is Hermitian;

(c) if all eigenvalues of \( A \) have unit absolute value then \( A \) is unitary.

We note that all of these implications are actually “if and only if,” as we shall demonstrate (Theorem 12.4.14).

Proposition 12.4.6. \( A \in M_n(\mathbb{C}) \) has an orthonormal eigenbasis if and only if \( A \) is unitarily similar to a diagonal matrix.

We now show (Cor. 12.4.8) that this condition is equivalent to \( A \) being normal.

Lemma 12.4.7. Let \( A, B \in M_n(\mathbb{C}) \) with \( A \sim_u B \). Then \( A \) is normal if and only if \( B \) is normal.

Corollary 12.4.8. If \( A \) is unitarily similar to a diagonal matrix, then \( A \) is normal.

Unitary similarity is a powerful tool to study matrices, owing largely to the following theorem.

Theorem 12.4.9 (Schur). Every matrix \( A \in M_n(\mathbb{C}) \) is unitarily similar to a triangular matrix.

Definition 12.4.10 (Dense subset). A subset \( S \subseteq \mathbb{F}^{k \times n} \) is dense in \( \mathbb{F}^{k \times n} \) if for every \( A \in S \) and every \( \varepsilon > 0 \), there exists \( B \in S \) such that every entry of the matrix \( A - B \) has absolute value less than \( \varepsilon \).

Proposition 12.4.11. Diagonalizable matrices are dense in \( M_n(\mathbb{C}) \).

Exercise 12.4.12. Complete the proof of the Cayley-Hamilton Theorem over \( \mathbb{C} \) (Theorem 8.5.3) using Prop. 12.4.11.

We now generalize the Spectral Theorem to normal matrices.
Theorem 12.4.13 (Complex Spectral Theorem). Let $A \in M_n(\mathbb{C})$. Then $A$ has an orthonormal eigenbasis if and only if $A$ is normal.

Before giving the proof, let us restate the theorem.

Theorem 12.4.14 (Complex Spectral Theorem, restated). Let $A \in M_n(\mathbb{C})$. Then $A$ is unitarily similar to a diagonal matrix if and only if $A$ is normal.

Exercise 12.4.15. Prove that Theorems 12.4.13 and 12.4.14 are equivalent.

Exercise 12.4.16. Infer Theorem 12.4.14 from Schur’s Theorem (Theorem 12.4.9) via the following lemma.

Lemma 12.4.17. If a triangular matrix is normal, then it is diagonal.

Theorem 12.4.18 (Real version of Schur’s Theorem). If $A \in M_n(\mathbb{R})$ and all eigenvalues of $A$ are real, then $A$ is orthogonally similar to an upper triangular matrix.

Exercise 12.4.19 (Third proof of the real Spectral Theorem). Infer the Spectral Theorem for real symmetric matrices (Theorem 10.1.1) from the complex Spectral Theorem and the fact, separately proved, that all eigenvalues of a symmetric matrix are real (Theorem 12.3.4).

12.5 Additional exercises

Exercise 12.5.1. Find an $n \times n$ unitary circulant matrix (Def. 2.5.12) with no zero entries.

Definition 12.5.2 (Discrete Fourier Transform matrix). The Discrete Fourier Transform (DFT) matrix is the $n \times n$ Vandermonde matrix (Def. 2.5.9) $F$ generated by $1, \omega, \omega^2, \ldots, \omega^{n-1}$, where $\omega$ is a primitive $n$th root of unity.

Exercise 12.5.3. Let $F$ be the $n \times n$ DFT matrix. Prove that $\frac{1}{\sqrt{n}} F$ is unitary.

Exercise 12.5.4. Which circulant matrices are unitary?
Chapter 13

(ℂ, ℜ) Matrix Norms

13.1 (ℜ) Operator norm

Definition 13.1.1 (Operator norm). The operator norm of a matrix \( A \in ℜ^{k \times n} \) is defined to be

\[
\| A \| = \max_{x \in ℜ^n, x \neq 0} \frac{\| Ax \|}{\| x \|}
\]

where the \( \| \cdot \| \) notation on the right-hand side represents the Euclidean norm.

Proposition 13.1.2. Let \( A \in ℜ^{k \times n} \). Then \( \| A \| \) exists.

Proposition 13.1.3. Let \( A \in ℜ^{k \times n} \) and \( \lambda \in ℜ \). Then \( \| \lambda A \| = |\lambda| \| A \| \).

Proposition 13.1.4 (Triangle inequality). Let \( A, B \in ℜ^{k \times n} \). Then \( \| A + B \| \leq \| A \| + \| B \| \).

Proposition 13.1.5 (Submultiplicativity). Let \( A \in ℜ^{k \times n} \) and \( B \in ℜ^{n \times ℓ} \). Then \( \| AB \| \leq \| A \| \cdot \| B \| \).

Exercise 13.1.6. Let \( A = [a_1 | \cdots | a_n] \in ℜ^{k \times n} \). (The \( a_i \) are the columns of \( A \).) Show \( \| A \| \geq \| a_i \| \) for every \( i \).

Exercise 13.1.7. Let \( A = (a_{ij}) \in ℜ^{k \times n} \). Show that \( \| A \| \geq |a_{ij}| \) for every \( i \) and \( j \).

Proposition 13.1.8. If \( A \) is a orthogonal matrix then \( \| A \| = 1 \).

Proposition 13.1.9. Let \( A \in ℜ^{k \times n} \). Let \( S \in O(k) \) and \( T \in O(n) \) be orthogonal matrices. Then

\[
\| SA \| = \| A \| = \| AT \|.
\]

Proposition 13.1.10. Let \( A \) be a symmetric real matrix with eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_n \). Then \( \| A \| = \max |\lambda_i| \).

Exercise 13.1.11. Let \( A \in ℜ^{k \times n} \). Show that \( A^T A \) is positive semidefinite (Def. 11.3.6).

Exercise 13.1.12. Let \( A \in ℜ^{k \times n} \) and let \( A^T A \) have eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_n \). Show that \( \| A \| = \sqrt{\lambda_1} \).

Proposition 13.1.13. For all \( A \in ℜ^{k \times n} \), we have \( \| A^T \| = \| A \| \).


(a) Find a stochastic matrix (Def. 22.1.2) of norm greater than 1.

(b) Find an \( n \times n \) stochastic matrix of norm \( \sqrt{n} \).

(c) Show that an \( n \times n \) stochastic matrix cannot have norm greater than \( \sqrt{n} \).
Numerical exercise 13.1.15.

(a) Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Calculate $\|A\|$.

(b) Let $B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Calculate $\|B\|$.

13.2 (R) Frobenius norm

Definition 13.2.1 (Frobenius norm). Let $A = (a_{ij}) \in \mathbb{R}^{k \times n}$ be a matrix. The Frobenius norm of $A$, denoted $\|A\|_F$, is defined as

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}.$$  \hfill (13.3)

Proposition 13.2.2. Let $A \in \mathbb{R}^{k \times n}$. Then $\|A\|_F = \sqrt{\text{Tr}(A^T A)}$.

Proposition 13.2.3. Let $A \in \mathbb{R}^{k \times n}$ and $\lambda \in \mathbb{R}$. Then $\|\lambda A\|_F = |\lambda| \|A\|_F$.

Proposition 13.2.4 (Triangle inequality). Let $A, B \in \mathbb{R}^{k \times n}$. Then $\|A + B\|_F \leq \|A\|_F + \|B\|_F$.

Proposition 13.2.5 (Submultiplicativity). Let $A \in \mathbb{R}^{k \times n}$ and $B \in \mathbb{R}^{n \times \ell}$. Then $\|AB\|_F \leq \|A\|_F \cdot \|B\|_F$.

Proposition 13.2.6. If $A \in O(k)$ then $\|A\|_F = \sqrt{k}$.

Proposition 13.2.7. Let $A \in \mathbb{R}^{k \times n}$. Let $S \in O(k)$ and $T \in O(n)$ be orthogonal matrices. Then $\|SA\|_F = \|A\|_F = \|AT\|_F$.

Proposition 13.2.8. For all $A \in \mathbb{R}^{k \times n}$, we have $\|A^T\|_F = \|A\|_F$.

Proposition 13.2.9. Let $A \in \mathbb{R}^{k \times n}$. Then

$$\|A\| \leq \|A\|_F \leq \sqrt{n} \|A\|.$$  \hfill (13.5)

Exercise 13.2.10. Prove $\|A\| = \|A\|_F$ if and only if $\text{rk} \ A = 1$. Use the Singular Value Decomposition (Theorem 21.1.2).

Exercise 13.2.11. Let $A$ be a symmetric real matrix. Show that $\|A\|_F = \sqrt{n} \|A\|$ if and only if $A = \lambda R$ for some reflection matrix $R$.

13.3 (C) Complex Matrices

Exercise 13.3.1. Generalize the definition of the operator norm and statements 13.1.2-13.1.12 to $\mathbb{C}$.

Exercise 13.3.2. Generalize the definition of the Frobenius norm and statements 13.2.2-13.2.10 to $\mathbb{C}$.

Exercise 13.3.3. Let $A$ be a normal matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that $\|A\|_F = \sqrt{n} \|A\|$ if and only if $|\lambda_1| = \cdots = |\lambda_n|$.

Question 13.3.4. Is normality necessary?
Part II

Linear Algebra of
Vector Spaces
Introduction to Part II

TO BE WRITTEN.
Chapter 14

(ℤ) Preliminaries

14.1 Modular arithmetic

Notation 14.1.1. The integers {..., −2, −1, 0, 1, 2, ...} are denoted by ℤ. The set {1, 2, 3, ...} of positive integers is denoted by ℤ+.

Definition 14.1.2 (Sum of sets). Let \(A, B \subseteq ℤ\). Then \(A + B\) is the set

\[ A + B = \{a + b \mid a \in A, b \in B\}. \tag{14.1} \]

Definition 14.1.3 (Cartesian product). Let \(A\) and \(B\) be sets. Then the Cartesian product of \(A\) and \(B\) is the set \(A \times B\) defined by

\[ A \times B = \{(a, b) \mid a \in A, b \in B\}. \tag{14.2} \]

Notice that \((a, b)\) in the above definition is an ordered pair. In particular, if \(a_1, a_2 \in A\), then \((a_1, a_2)\) and \((a_2, a_1)\) are distinct elements of \(A \times A\).

Notation 14.1.4 (Cardinality). For a set \(A\) we denote the cardinality of \(A\) (the number of elements of \(A\)) by \(|A|\).

For instance, \(|\{4, 5, 6, 4\}| = 3.\]

Exercise 14.1.5. Let \(A, B \subseteq ℤ\) with \(|A| = n\) and \(|B| = m\). Show that

(a) \(m + n - 1 \leq |A + B| \leq mn\)
(b) \(|A + A| \leq \binom{n+1}{2}\)
(c) \(|A + A + A| \leq \binom{n+2}{3}\)
(d) \(|A \times B| = mn\)

Definition 14.1.6 (Divisibility). Let \(a, b \in ℤ\). Then \(a\) divides \(b\), or \(a\) is a divisor of \(b\) (notation: \(a \mid b\)) if there exists some \(c \in ℤ\) such that \(b = ac\).

Proposition 14.1.7. For all \(a \in ℤ\), \(1 \mid a\).

Proposition 14.1.8. Show that \(0 \mid 0\).

Notice that this does not violate any rules that we know about division by 0 because our definition of divisibility does not involve division.

Exercise 14.1.9. For which \(a\) do we have \(a \mid b\) for all \(b \in ℤ\)?

Exercise 14.1.10. For what \(b\) do we have \(a \mid b\) for all \(a \in ℤ\)?

Proposition 14.1.11. Let \(a, b, c \in ℤ\). If \(c \mid a\) and \(c \mid b\) then \(c \mid a + b\).

Proposition 14.1.12 (Transitivity of divisibility). Let \(a, b, c \in ℤ\). If \(a \mid b\) and \(b \mid c\), then \(a \mid c\).
**Definition 14.1.13 (Congruence modulo \( m \)).** Let \( a, b, m \in \mathbb{Z} \). Then \( a \) is congruent to \( b \) modulo \( m \) (written \( a \equiv b \pmod{m} \)) if \( m \mid a - b \).

**Exercise 14.1.14.** Let \( a, b \in \mathbb{Z} \). Show
(a) \( a \equiv b \pmod{1} \)
(b) \( a \equiv b \pmod{2} \) if and only if \( a \) and \( b \) are either both even or both odd
(c) \( a \equiv b \pmod{0} \) if and only if \( a = b \)

**Exercise 14.1.15.** Let \( a, b, c, d, m \in \mathbb{Z} \). Show that if \( a \equiv c \pmod{m} \) and \( b \equiv d \pmod{m} \), then
(a) \( a + b \equiv c + d \pmod{m} \)
(b) \( a - b \equiv c - d \pmod{m} \)
(c) \( ab \equiv cd \pmod{m} \)

**Exercise 14.1.16.** Let \( k \geq 0 \). Show that if \( a \equiv b \pmod{m} \), then \( a^k \equiv b^k \pmod{m} \).

**Example 14.1.17.** Calendar arithmetic is an application of modular arithmetic; for example, if August 3 is a Wednesday, then August 24 is also a Wednesday, because \( 3 \equiv 24 \pmod{7} \).

**Definition 14.1.18 (Binary relation).** A binary relation on the set \( A \) is a subset \( R \subseteq A \times A \). The relation \( R \) holds for the elements \( a, b \in A \) if \( (a, b) \in R \). In this case, we write \( a R b \) or \( R(a, b) \).

**Definition 14.1.19 (Equivalence relation).** Let \( \sim \) be a binary relation on a set \( A \). The relation \( \sim \) is said to be an equivalence relation if the following properties hold for all \( a, b, c \in A \).
(a) \( a \sim a \) (reflexivity)
(b) \( a \sim b \) if and only if \( b \sim a \) (symmetry)
(c) If \( a \sim b \) and \( b \sim c \), then \( a \sim c \) (transitivity)

**Proposition 14.1.20.** For all \( m \in \mathbb{Z} \), “congruence modulo \( m \)” is an equivalence relation.

**Definition 14.1.21 (Equivalence class).** Let \( A \) be a set with an equivalence relation \( \sim \), and let \( a \in A \). The equivalence class of \( a \) with respect to \( \sim \), denoted \([a]\), is the set
\[
[a] = \{ b \in A \mid a \sim b \} \tag{14.3}
\]
The equivalence classes of the equivalence relation “congruence modulo \( m \)” in \( \mathbb{Z} \) are called residue classes of \( \mathbb{Z} \) modulo \( m \).

**Proposition 14.1.22.** Let \( \sim \) be an equivalence relation on a set \( A \). Then our choice of representatives for equivalence classes does not matter; that is, if \( a \sim b \) then \([a] = [b]\).

**Exercise 14.1.23.** For each \( m \in \mathbb{Z} \), how many residue classes modulo \( m \) are there?

**Definition 14.1.24 (Sum of residue classes).** Let \([a]\) and \([b]\) be residue classes of \( \mathbb{Z} \) modulo \( m \). Then their sum is
\[
[a] + [b] = [a + b] \tag{14.4}
\]

**Definition 14.1.25 (Product of residue classes).** Let \([a]\) and \([b]\) be residue classes of \( \mathbb{Z} \) modulo \( m \). Then their product is
\[
[a] \cdot [b] = [a \cdot b] \tag{14.5}
\]
**Exercise 14.1.26.** Show that the sum and product of residue classes are well defined, that is, that they do not depend on our choice of representative for each residue class.

**Convention 14.1.27.** The residue classes of \( \mathbb{Z} \) modulo \( m \) are often represented by the integers \( 0, 1, \ldots, m - 1 \).

**Proposition 14.1.28.** Let \( \mathbb{Z}_m \) denote the set of residue classes modulo \( m \) with the operations of addition and multiplication as defined above. Verify the following properties of these operations.

(a) For all \( a, b \in \mathbb{Z}_m \), there exist unique \( a + b \in \mathbb{Z}_m \) and \( a \cdot b \in \mathbb{Z}_m \)

(b) For all \( a, b, c \in \mathbb{Z}_m \) we have \((a + b) + c = a + (b + c)\) and \((a \cdot b) \cdot c = a \cdot (b \cdot c)\) (associativity)

(c) There exists an additive identity element in \( \mathbb{Z}_m \), which we denote by 0, such that for all \( a \in \mathbb{Z}_m \) we have \( 0 + a = a + 0 = a \) (additive identity)

(d) There exists a multiplicative identity element in \( \mathbb{Z}_m \), which we denote by 1, such that for all \( a \in \mathbb{Z}_m \) we have \( 1 \cdot a = a \cdot 1 = a \) (multiplicative identity)

(e) For each \( a \in \mathbb{Z}_m \), there exists an element \( -a \in \mathbb{Z}_m \) such that \( a + (-a) = 0 \) (additive inverse)

(f) For all \( a, b \in \mathbb{Z}_m \), we have \( a + b = b + a \) and \( a \cdot b = b \cdot a \) (commutativity)

(g) For all \( a, b, c \in \mathbb{Z}_m \), \( a \cdot (b + c) = a \cdot b + a \cdot c \) (distributivity)

Further, if \( p \) is a prime number, then

(h) For all \( a \in \mathbb{Z}_p \), if \( a \neq 0 \) then there exists an element \( a^{-1} \in \mathbb{Z}_p \) such that \( a \cdot a^{-1} = 1 \) (multiplicative inverse)

### 14.2 Abelian groups

A group is a set with a binary operation satisfying certain axioms. The set \( \mathbb{Z} \) with the operation of addition is an example of group. A class of further examples are the sets \( \mathbb{Z}_m \) of residue classes modulo \( m \), with addition being the operation.

We now generalize this notion to include a much larger collection of algebraic structures.

**Definition 14.2.1 (Group).** A group is a set \( G \) along with a binary operation \( \circ \) that satisfies the following axioms.

(a) For all \( a, b \in G \), there exists a unique element \( a \circ b \in G \)

(b) For all \( a, b, c \in G \), \((a \circ b) \circ c = a \circ (b \circ c)\) (associativity)

(c) There exists an identity element \( e \in G \) such that for all \( a \in G \), \( e \circ a = a \circ e = a \) (identity element)

(d) For each \( a \in G \), there exists \( b \in G \) such that \( a \circ b = b \circ a = e \) (inverses)
Proposition 14.2.2. The identity element of $G$ is unique. For every $a \in G$, the inverse of $a$ is unique.

The inverse of an element $a$ is often written as $a^{-1}$, and we often write $ab$ instead of $a \circ b$. The set $G$ along with the binary operation $\circ$ is the group $(G, \circ)$. We often omit $\circ$ and we refer to $G$ as the group when the binary operation is clear from context. Groups satisfying the additional axiom

$$(e) \quad a \circ b = b \circ a \text{ for all } a, b \in G \text{ (commutativity)}$$

are called abelian groups.

When discussing abelian groups, we typically use additive notation (that is, we write the binary operation $\circ$ as $+$), and we write the identity as 0 and the inverse of an element $a$ as $-a$.

Definition 14.2.3 (Order of a group). Let $G$ be a group. The order of $G$ is its cardinality, $|G|$.

Observe that Prop. 14.1.28 asserted that $(\mathbb{Z}_m, +)$ is a group for $m \geq 0$ and that $(\mathbb{Z}_p, \cdot)$ is a group when $p$ is prime. We now present more examples of groups.

Examples 14.2.4. Show that the following are abelian groups.

(a) The integers with addition, $(\mathbb{Z}, +)$

(b) The real numbers with addition, $(\mathbb{R}, +)$

(c) The real numbers except for 0 with multiplication, $(\mathbb{R}^\times, \times)$

(d) For any set $\Omega$, the set $\mathbb{R}^\Omega$ of real functions $f : \Omega \to \mathbb{R}$ with pointwise addition, $(\mathbb{R}^\Omega, +)$, where pointwise addition is defined by the rules $(f + g)(\omega) = f(\omega) + g(\omega)$ for all $\omega \in \Omega$.

(e) The complex $n$-th roots of unity, that is, the set

$$\{\omega \in \mathbb{C} \mid \omega^n = 1\}$$

with multiplication.

In particular, it follows from (e) that there exist abelian groups of every finite order.

Definition 14.2.5. We define the sum $\sum_{i=1}^{k} a_i$ by induction. For the base case of $k = 1$,

$$\sum_{i=1}^{1} a_i = a_i,$$

and for $k > 1$,

$$\sum_{i=1}^{k} a_i = a_k + \sum_{i=1}^{k-1} a_i . \quad (14.6)$$

Convention 14.2.6 (Empty sum). The empty sum is equal to 0, that is,

$$\sum_{i=1}^{0} a_i = 0 . \quad (14.7)$$

Convention 14.2.7 (Empty product). By convention, the empty product $\prod_{a \in \emptyset} a$ is equal to 1.

In particular, $0^0 = 1$ and $0! = 1$. 


Exercise 14.2.8. Let $G$ be an abelian group and let $A, B \subseteq G$ with $|A| = n$ and $B = m$. Show that

(a) $|A + B| \leq mn$

(b) $|A + A| \leq \left(\frac{n+1}{2}\right)$

(c) $|A + A + A| \leq \left(\frac{n+2}{3}\right)$

Definition 14.2.9 (Subgroup). Let $G$ be a group. If $H \subseteq G$ is nonempty and is a group under the same operation as $G$, we say that $H$ is a subgroup of $G$, denoted $H \leq G$.

Proposition 14.2.10. Let $G$ be a group and $H \subseteq G$ be a subset. $H$ is a subgroup of $G$ if and only if

(a) The identity element of $G$ is a member of $H$. (In additive notation, $0 \in H$.)

(b) $H$ is closed under the operation: if $a, b \in H$ then $ab \in H$. (In additive notation, $a + b \in H$.)

(c) $H$ is closed under inversion: if $a \in H$ then $a^{-1} \in H$. (In additive notation, $-a \in H$.)

Proposition 14.2.11. The relation $\leq$ is transitive, that is, if $K \leq H$ and $H \leq G$, then $K \leq G$.

Proposition 14.2.12. The intersection of any collection of subgroups of a group $G$ is itself a subgroup of $G$.

Exercise 14.2.13. Let $G$ be a group and let $H, K \in G$. Then $H \cup K \leq G$ if and only if $H \subseteq K$ or $K \subseteq H$.

Proposition 14.2.14. Let $G$ be a group and let $H \leq G$. Then

(a) The identity of $H$ is the same as the identity of $G$.

(b) Let $a \in H$. The inverse of $a$ in $H$ is the same as the inverse of $a$ in $G$.

Notation 14.2.15. Let $G$ be an abelian group, written additively. Let $H, K \subseteq G$. We write $-H$ for the set $-H = \{-h \mid h \in H\}$, and $H - K$ for the set

$$H - K = H + (-K) = \{h - k \mid h \in H, k \in K\}$$

(14.8)

Proposition 14.2.16. Let $G$ be an abelian group, written additively. Let $H \subseteq G$. Then $H \leq G$ if and only if

(a) $0 \in H$

(b) $-H \subseteq H$ (closed under additive inverses)

(c) $H + H \subseteq H$ (closed under addition)

Proposition 14.2.17. Let $G$ be an abelian group, written additively. Let $H \subseteq G$. Then $H \leq G$ if and only if $H \neq \emptyset$ and $H - H \subseteq H$ (that is, $H$ is closed under subtraction).
14.3 Fields

Definition 14.3.1 (Field). A field $\mathbb{F}$ is a set with two operations, addition and multiplication, satisfying the following axioms.

(a) $(\mathbb{F}, +)$ is an abelian group.

(b) $(\mathbb{F}^\times, \cdot)$ is an abelian group, where $\mathbb{F}^\times = \mathbb{F} \setminus \{0\}$.

(c) Distributivity holds: $a (b + c) = ab + ac$.

Proposition 14.3.2. For all $a \in \mathbb{F}$ we have $0 \cdot a = 0$ and $1 \cdot a = a$.

Exercise 14.3.3. (a) $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers), $\mathbb{C}$ (complex numbers) are examples of fields.

(b) $\mathbb{Z}$ is not a field.

(c) For $p$ a prime number, we define $\mathbb{F}_p = \mathbb{Z}_p$ the set of residue classes modulo $p$ with addition and multiplication as the operations. $\mathbb{F}_p$ is a field.

(d) $\mathbb{Z}_m$ is a field if and only if $m$ is a prime number.

The field $\mathbb{F}_p$ has $p$ elements; there are finite fields other than $\mathbb{F}_p$ as well. The number of elements of a field is referred to as the order of the field. Fields were invented by Évariste Galois (1811-1832) (along with groups and the notion of abstract algebra). Galois showed that (a) the order of every finite field is a prime power and (b) for every prime power $q$ there exists a unique field $\mathbb{F}_q$ of order $q$. Note that $\mathbb{F}_q \neq \mathbb{Z}_q$ unless $q$ is a prime.

Exercise 14.3.4. (a) Prove that the set $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a field.

(b) Prove that the set $\mathbb{Q}[\sqrt[3]{2}] = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in \mathbb{Q}\}$ is a field.

Exercise 14.3.5. Let $p$ be a prime. Consider the set $\mathbb{F}_p[i]$ of formal expressions of the form $a + bi$ ($a, b \in \mathbb{F}_p$) where multiplication is performed by observing the rule $i^2 = -1$. Determine, for what primes $p$ is $\mathbb{F}_p[i]$ a field. (Experiment, find a pattern, prove.) This exercise will give you an infinite number of fields of order $p^2$.

14.4 Polynomials

In Section 8.3 we developed a basic theory of polynomials. In that section, however, we considered polynomials as functions. We now develop a more formal theory of polynomials, first considering a polynomial $f$ as a formal expression whose coefficients are taken from a field $\mathbb{F}$ of scalars, rather than as a function $f : \mathbb{F} \to \mathbb{F}$.

Definition 14.4.1 (Polynomial). A polynomial over the field $\mathbb{F}$ is an expression of the form

$$f = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n \quad (14.9)$$

where the coefficients $\alpha_i$ are scalars (elements of $\mathbb{F}$), and $t$ is a symbol. The set of all polynomials over $\mathbb{F}$ is denoted $\mathbb{F}[t]$. Two expressions,
CHAPTER 14. (\( \mathbb{F} \)) PRELIMINARIES

(14.9) and
\[
g = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_m t^m
\]  
(14.10)

\( g \) define the same polynomial if they only differ in leading zero coefficients, i.e., there is some \( k \) for which \( \alpha_0 = \beta_0, \ldots, \alpha_k = \beta_k \), and all coefficients \( \alpha_j, \beta_j \) are zero for \( j > k \). We may omit any terms with zero coefficient, e.g.,
\[
3 + 0t + 2t^2 + 0t^3 = 3 + 2t^2
\]  
(14.11)

Definition 14.4.2 (Zero polynomial). The polynomial which has all coefficients equal to zero is called the zero polynomial and is denoted by \( 0 \).

Definition 14.4.3 (Leading term). The leading term of a polynomial \( f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n \) is the term corresponding to the highest power of \( t \) with a nonzero coefficient, that is, the term \( \alpha_k t^k \) where \( \alpha_k \neq 0 \) and \( \alpha_j = 0 \) for all \( j > k \). The zero polynomial does not have a leading term.

Definition 14.4.4 (Leading coefficient). The leading coefficient of a polynomial \( f \) is the coefficient of the leading term of \( f \).

For example, the leading term of the polynomial \( 3 + 2t^2 + 5t^7 \) is \( 5t^7 \) and the leading coefficient is 5.

Definition 14.4.5 (Monic polynomial). A polynomial is monic if its leading coefficient is 1.

Definition 14.4.6 (Degree of a polynomial). The degree of a polynomial \( f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n \), denoted \( \deg f \), is the exponent of its leading term.

For example, \( \deg (3 + 2t^2 + 5t^7) = 7 \).

Convention 14.4.7. The zero polynomial has degree \( -\infty \).

Exercise 14.4.8. Which polynomials have degree 0?

Notation 14.4.9. We denote the set of polynomials of degree at most \( n \) over \( \mathbb{F} \) by \( P_n[\mathbb{F}] \).

Definition 14.4.10 (Sum and difference of polynomials). Let \( f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n \) and \( g = \beta_0 + \beta_1 t + \cdots + \beta_n t^n \) be polynomials. Then the sum of \( f \) and \( g \) is defined as
\[
f + g = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) t + \cdots + (\alpha_n + \beta_n) t^n
\]  
(14.12)

and the difference \( f - g \) is defined as
\[
f - g = (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) t + \cdots + (\alpha_n - \beta_n) t^n
\]  
(14.13)

Note that \( f \) and \( g \) need not be of the same degree; we can add on leading zeros if necessary.

Numerical exercise 14.4.11. Let
\[
f = 2t + t^2
\]
\[
g = 3 + 3t^2 + 3t^3
\]
\[
h = 5 + t^3 + t^4
\]

Compute the polynomials
(a) \( e_1 = f - g \)
(b) \( e_2 = g - h \)
(c) \( e_3 = h - f \)

Self-check: verify that \( e_1 + e_2 + e_3 = 0 \).
14.4. POLYNOMIALS

**Proposition 14.4.12.** Addition of polynomials is (a) commutative and (b) associative, that is, if \( f, g, h \in \mathbb{F}[t] \) then

(a) \( f + g = g + f \)

(b) \( f + (g + h) = (f + g) + h \)

**Definition 14.4.13 (Multiplication of polynomials).** Let \( f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n \) and \( g = \beta_0 + \beta_1 t + \cdots + \beta_m t^m \) be polynomials. Then the **product** of \( f \) and \( g \) is defined as

\[
f \cdot g = \sum_{i=0}^{n+m} \left( \sum_{j+k=i} \alpha_j \beta_k \right) t^i . \tag{14.14}\]

**Numerical exercise 14.4.14.** Let \( f, g, \) and \( h \) be as in Example 14.4.11. Compute

(a) \( e_4 = f \cdot g \)

(b) \( e_5 = f \cdot h \)

(c) \( e_6 = f \cdot (g + h) \)

**Self-check:** verify that \( e_4 + e_5 = e_6 \).

**Proposition 14.4.15.** Let \( f, g \in \mathbb{F}[t] \). Then

(a) \( \deg(f + g) \leq \max\{\deg f, \deg g\} \),

(b) \( \deg(fg) = \deg f + \deg g \).

Note that both of these statements hold even if one of the polynomials is the zero polynomial.

**Notation 14.4.16 (Set of functions).** Let \( A \) and \( B \) be sets. The set of functions \( f : B \to A \) is denoted by \( A^B \). Here \( B \) is the **domain** of the functions \( f \) in question and \( A \) is the **target set** (into which \( f \) maps \( B \)).

The following exercise explains the reason for this notation.

**Exercise 14.4.17.** Let \( A \) and \( B \) be finite sets. Show that \( |A^B| = |A|^{|B|} \).

**Definition 14.4.18 (Substitution).** For \( \zeta \in \mathbb{F} \) and \( f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n \in \mathbb{F}[t] \), we set

\[ f(\zeta) = \alpha_0 + \alpha_1 \zeta + \cdots + \alpha_n \zeta^n \in \mathbb{F} \quad (14.15) \]

The substitution \( t \mapsto \zeta \) defines a mapping \( \mathbb{F}[t] \to \mathbb{F} \) which assigns the value \( f(\zeta) \) to \( f \). We denote the \( \mathbb{F} \to \mathbb{F} \) function \( \zeta \mapsto f(\zeta) \) by \( \overline{f} \) and call \( \overline{f} \) a **polynomial function**. So \( \overline{f} \) is a function while \( f \) is a formal expression. If \( A \) is an \( n \times n \) matrix, then we define

\[ f(A) = \alpha_0 I + \alpha_1 A + \cdots + \alpha_n A^n \in M_n(\mathbb{F}) \] \quad (14.16)

**Definition 14.4.19 (Divisibility of polynomials).** Let \( f, g \in \mathbb{F}[t] \). We say that \( g \) **divides** \( f \), or \( f \) is **divisible** by \( g \), written \( g \mid f \), if there exists a polynomial \( h \in \mathbb{F}[t] \) such that \( f = gh \). In this case we say that \( g \) is a **divisor** of \( f \) and \( f \) is a **multiple** of \( g \).

**Notation 14.4.20 (Divisors of a polynomial).** Let \( f \in \mathbb{F}[t] \). We denote by \( \text{Div}(f) \) the set of all divisors of \( f \). We denote by \( \text{Div}(f, g) \) the set of polynomials which divide both \( f \) and \( g \), that is, \( \text{Div}(f, g) = \text{Div}(f) \cap \text{Div}(g) \).

**Theorem 14.4.21 (Division Theorem).** Let \( f, g \in \mathbb{F}[t] \) where \( g \) is not be the zero polynomial. Then there exist unique polynomials \( q \) and \( r \) such that

\[ f = qg + r \quad \tag{14.17} \]

and \( \deg r < \deg g \). \( \diamond \)
Exercise 14.4.22. Let \( f \in \mathbb{F}[t] \) and \( \zeta \in \mathbb{F} \).

(a) Show that there exist \( q \in \mathbb{F}[t] \) and \( \xi \in \mathbb{F} \) such that
\[
 f = (t - \zeta)q + \xi
\]

(b) Prove that \( \xi = f(\zeta) \)

Exercise 14.4.23. Let \( f \in \mathbb{F}[t] \).

(a) Show that if \( \mathbb{F} \) is an infinite field, then \( f = 0 \) if and only if \( f = 0 \).

(b) Let \( \mathbb{F} \) be a finite field of order \( q \).

(b1) Prove that there exists \( f \neq 0 \) such that \( \overline{f} = 0 \).

(b2) Show that if \( \deg f < q \), then \( f = 0 \) if and only if \( \overline{f} = 0 \). Do not use (b3) to solve this; your solution should be just one line, based on a very simple formula defining \( f \).

(b3) \(^{(*)}\) (Fermat’s Little Theorem, generalized) Show that if \( |\mathbb{F}| = q \) and \( f = t^q - t \), then \( \overline{f} = 0 \).

Definition 14.4.24 (Ideal). Let \( I \subseteq \mathbb{F}[t] \). Then \( I \) is an ideal if the following three conditions hold.

(a) \( 0 \in I \)

(b) \( I \) is closed under addition

(c) If \( f \in I \) and \( g \in \mathbb{F}[t] \) then \( fg \in I \)

Notation 14.4.25. Let \( f \in \mathbb{F}[t] \). We denote by \( (f) \) the set of all multiples of \( f \), i.e.,
\[
 (f) = \{ fg \mid g \in \mathbb{F}[t] \} .
\]

Proposition 14.4.26. Let \( f, g \in \mathbb{F}[t] \). Then \( (f) \subseteq (g) \) if and only if \( g \mid f \).

Definition 14.4.27 (Principal ideal). Let \( f \in \mathbb{F}[t] \). The set \( (f) \) is called the principal ideal generated by \( f \), and \( f \) is said to be a generator of this ideal.

Exercise 14.4.28. For \( f \in \mathbb{F}[t] \), verify that \( (f) \) is indeed an ideal of \( \mathbb{F}[t] \).

Theorem 14.4.29 (Every ideal is principal). Every ideal of \( \mathbb{F}[t] \) is principal.

Proposition 14.4.30. Let \( f, g \in \mathbb{F}[t] \). Then \( (f) = (g) \) if and only if there exists \( \zeta \neq 0 \in \mathbb{F} \) such that \( g = \alpha f \).

Theorem 14.4.29 will be our tool to prove the existence of greatest common divisors of polynomials.

Definition 14.4.31 (Greatest common divisor). Let \( f_1, \ldots, f_k, g \in \mathbb{F}[t] \). We say that \( g \) is a greatest common divisor (gcd) of \( f_1, \ldots, f_k \) if

(a) \( g \) is a common divisor of the \( f_i \), i.e., \( g \mid f_i \) for all \( i \),

(b) \( g \) is a common multiple of all common divisors of the \( f_i \), i.e., for all \( e \in \mathbb{F}[t] \), if \( e \mid f_i \) for all \( i \), then \( e \mid g \).

Proposition 14.4.32. Let \( f_1, \ldots, f_k, d, d_1, d_2 \in \mathbb{F}[t] \).
(a) Let \( \zeta \) be a nonzero scalar. If \( d \) is a gcd of \( f_1, \ldots, f_k \), then \( \zeta d \) is also a gcd of \( f_1, \ldots, f_k \).

(b) If \( d_1 \) and \( d_2 \) are both gcds of \( f_1, \ldots, f_k \), then there exists \( \zeta \in F^\times = F \setminus \{0\} \) such that \( d_2 = \zeta d_1 \).

**Exercise 14.4.33.** Let \( f_1, \ldots, f_k \in F[t] \). Show that \( \gcd(f_1, \ldots, f_k) = 0 \) if and only if \( f_1 = \cdots = f_k = 0 \).

**Proposition 14.4.34.** Let \( f_1, \ldots, f_k \in F[t] \) and suppose not all of the \( f_i \) are 0. Then among all of the greatest common divisors of \( f_1, \ldots, f_k \), there is a unique monic polynomial.

For the sake of uniqueness of the gcd notation, we write \( d = \gcd(f_1, \ldots, f_k) \) if, in addition to (a) and (b),

(c) \( d \) is monic or \( d = 0 \).

**Theorem 14.4.35 (Existence of gcd).** Let \( f_1, \ldots, f_k \in F[t] \). Then \( \gcd(f_1, \ldots, f_k) \) exists and, moreover, that there exist polynomials \( g_1, \ldots, g_k \) such that

\[
\gcd(f_1, \ldots, f_k) = \sum_{i=1}^{k} f_i g_i \quad (14.18)
\]

**Lemma 14.4.36 (Euclid’s Lemma).** Let \( f, g, h \in F[t] \). Then \( \Div(f, g) = \Div(f - gh, g) \).

**Theorem 14.4.37.** Let \( f_1, \ldots, f_k, g \in F[t] \). Then \( g \) is a gcd of \( f_1, \ldots, f_k \) if and only if \( \Div(f_1, \ldots, f_k) = \Div(g) \).

**Exercise 14.4.38.** Let \( f \in F[t] \). Show that \( \Div(f, 0) = \Div(f) \).

By applying the above theorem and Euclid’s Lemma, we arrive at *Euclid’s Algorithm* for determining the gcd of polynomials.

**Theorem 14.4.39 (Euclid’s Algorithm).** The following algorithm can be used to determine the greatest common divisor of two polynomials \( f_0 \) and \( g_0 \). In each iteration of the while loop, \( r = f - g \cdot q \), where \( r \) and \( q \) are the polynomials guaranteed by the Division Theorem. In particular, \( \deg r < \deg g \).

\[
f \leftarrow f_0 \\
g \leftarrow g_0 \\
\text{while } g \neq 0 \text{ do} \\
\quad \text{Find } q \text{ and } r \text{ by the Division Theorem} \\
\quad f \leftarrow g \\
\quad g \leftarrow r \\
\text{end while} \\
\text{return } f
\]

The following example provides a demonstration of Euclid’s algorithm.

**Example 14.4.40.** Let \( f = t^5 + 2t^4 - 3t^3 + \)
CHAPTER 14. \((\mathbb{F})\) PRELIMINARIES

\[ t^2 - 5t + 4 \text{ and let } g = t^2 - 1. \text{ Then} \]

\[ \text{Div}(f, g) = \text{Div} \left( f - (t^3 + 2t^2 - 2t + 3) g, g \right) \]
\[ = \text{Div} \left( t^2 - 1, -7t - 7 \right) \]
\[ = \text{Div} \left( t^2 - 1 + \frac{t}{7}(-7t + 7), -7t + 7 \right) \]
\[ = \text{Div} \left( t - 1, -7t + 7 \right) \]
\[ = \text{Div}(0, -7t + 7) \]
\[ = \text{Div}(-7t + 7) \]

Thus \(-7t + 7\) is a gcd of \(f\) and \(g\), and we can multiply by \(-\frac{1}{7}\) to get a monic polynomial. In particular, \(t - 1\) is the gcd of \(f\) and \(g\), and we may write \(\gcd(f, g) = t - 1\).

**Exercise 14.4.41.** Let

\[
\begin{align*}
f_1 &= t^2 + t - 2 \\
f_2 &= t^2 + 3t + 2 \\
f_3 &= t^3 - 1 \\
f_4 &= t^4 - t^2 - 2t - 1
\end{align*}
\]

Determine the following greatest common divisors.

(a) \(\gcd(f_1, f_2)\)

(b) \(\gcd(f_1, f_3)\)

(c) \(\gcd(f_1, f_3, f_4)\)

(d) \(\gcd(f_1, f_2, f_3, f_4)\)

**Proposition 14.4.42.** Let \(f, g, h \in \mathbb{F}[t]\). Then \(\gcd(fg, fh) = fd\), where \(d = \gcd(g, h)\).

**Proposition 14.4.43.** If \(f | gh\) and \(\gcd(f, g) = 1\), then \(f | h\).

**Exercise 14.4.44.** Determine \(\gcd(f, f')\) where \(f = t^n + t + 1\) (over \(\mathbb{R}\)).

**Exercise 14.4.45.** Determine \(\gcd(t^n - 1, t^2 + t + 1)\).

Let \(f \in \mathbb{F}[t]\) and let \(\mathbb{F}\) be a subfield of the field \(\mathbb{G}\). Then we can also view \(f\) as a polynomial over \(\mathbb{G}\). However, this changes the notion of linear combinations and therefore in principle it could affect the gcd of two polynomials. We shall see that this is not the case, but in order to be able to reason about this question, we temporarily use the notation \(\gcd_\mathbb{F}(f)\) and \(\gcd_\mathbb{G}(f)\) to denote the gcd of \(f\) with respect to the corresponding fields.

**Exercise 14.4.46 (Insensitivity of gcd to field extensions).** Let \(\mathbb{F}\) be a subfield of \(\mathbb{G}\), and let \(f, g \in \mathbb{F}[t]\). Then

\[ \gcd_\mathbb{F}(f, g) = \gcd_\mathbb{G}(f, g). \]

**Definition 14.4.47 (Irreducible polynomial).** A polynomial \(f \in \mathbb{F}[t]\) is **irreducible** over \(\mathbb{F}\) if \(\deg f \geq 1\) and for all \(g, h \in \mathbb{F}[t]\), if \(f = gh\) then either \(\deg g = 0\) or \(\deg h = 0\).

We shall give examples of irreducible polynomials over various fields in Examples ??

**Proposition 14.4.48.** If \(f \in \mathbb{F}[t]\) is irreducible and \(\zeta\) is a nonzero scalar, then \(\zeta f\) is irreducible.
14.4. POLYNOMIALS

Proposition 14.4.49. Let \( f \in \mathbb{F}[t] \) with \( \deg f = 1 \). Then \( f \) is irreducible.

Proposition 14.4.50. Let \( f \) be an irreducible polynomial, and let \( f \mid gh \). Then either \( f \mid g \) or \( f \mid h \).

Proposition 14.4.51. Every nonzero polynomial is a product of irreducible polynomials.

Theorem 14.4.52 (Unique Factorization). Every polynomial in \( \mathbb{F}[t] \) can be uniquely written as the product of irreducible polynomials over \( \mathbb{F} \).

Uniqueness holds up to the order of the factors and scalar multiples, i.e., if \( f_1 \cdots f_k = g_1 \cdots g_\ell \) where the \( f_i \) and \( g_j \) are irreducible, then \( k = \ell \) and there exists a permutation (bijection) \( \sigma : \{1, \ldots, k\} \to \{1, \ldots, k\} \) and nonzero scalars \( \alpha_i \in \mathbb{F} \) such that \( f_i = \alpha_i g_{\sigma(i)} \).

Definition 14.4.53 (Root of a polynomial). Let \( f \in \mathbb{F}[t] \). We say that \( \zeta \in \mathbb{F} \) is a root of \( f \) if \( f(\zeta) = 0 \).

Proposition 14.4.54. Let \( \zeta \in \mathbb{F} \) and let \( f \in \mathbb{F}[t] \). Then \( t - \zeta \mid f - f(\zeta) \).

Corollary 14.4.55. Let \( \zeta \in \mathbb{F} \) and \( f \in \mathbb{F}[t] \). Then \( \zeta \) is a root of \( f \) if and only if \( (t - \zeta) \mid f \).

Definition 14.4.56 (Multiplicity). The multiplicity of a root \( \zeta \) of a polynomial \( f \in \mathbb{F}[t] \) is the largest \( k \) for which \( (t - \zeta)^k \mid f \).

Exercise 14.4.57. Let \( f \in \mathbb{R}[t] \). Show that \( f(\sqrt{-1}) = 0 \) if and only if \( t^2 + 1 \mid f \).

Proposition 14.4.58. Let \( f \) be a polynomial of degree \( n \). Then \( f \) has at most \( n \) roots (counting multiplicity).

Theorem 14.4.59 (Fundamental Theorem of Algebra). Let \( f \in \mathbb{C}[t] \). If \( \deg f \geq 1 \), then \( f \) has a complex root, i.e., there exists \( \zeta \in \mathbb{C} \) such that \( f(\zeta) = 0 \).

Proposition 14.4.60. If \( f \in \mathbb{C}[t] \), then \( f \) is irreducible if and only if \( \deg f = 1 \).

Proposition 14.4.61. If \( f \in \mathbb{C}[t] \) and \( \deg f = k \geq 1 \), then \( f \) can be written as
\[
f = \alpha_k \prod_{i=1}^{k} (t - \zeta_i) \quad (14.23)
\]
where \( \alpha_k \) is the leading coefficient of \( f \) and the \( \zeta_i \) are complex numbers.

Proposition 14.4.62. Let \( f \in \mathbb{P}_2[\mathbb{R}] \) be given by \( f = at^2 + bt + c \) with \( a \neq 0 \). Then \( f \) is irreducible over \( \mathbb{R} \) if and only if \( b^2 - 4ac < 0 \).

Exercise 14.4.63. Let \( f \in \mathbb{F}[t] \).

(a) If \( f \) has a root in \( \mathbb{F} \) and \( \deg f \geq 2 \), then \( f \) is reducible.

(b) Find a reducible polynomial over \( \mathbb{R} \) that has no real root.

Proposition 14.4.64. Let \( f \in \mathbb{R}[t] \) be of odd degree. Then \( f \) has a real root.

Proposition 14.4.65. Let \( f \in \mathbb{R}[t] \) and \( \zeta \in \mathbb{C} \). Then \( f(\overline{\zeta}) = \overline{f(\zeta)} \). Conclude that if \( \zeta \) is a complex root of \( f \), then so is \( \overline{\zeta} \).
Exercise 14.4.66. Let \( f \in \mathbb{R}[t] \). Show that if \( f \) is irreducible over \( \mathbb{R} \), then \( \deg f \leq 2 \).

Exercise 14.4.67. Let \( f \in \mathbb{R}[t] \) and \( f \neq 0 \). Show that \( f \) can be written as \( f = \prod g_i \) where each \( g_i \) has degree 1 or 2.

Definition 14.4.68 (Formal derivative of a polynomial). Let \( f = \sum_{i=0}^{n} \alpha_i t^i \). Then the formal derivative of \( f \) is defined to be

\[
 f' = \sum_{k=1}^{n} k\alpha_k t^{k-1} \quad (14.24)
\]

That is,

\[
 f' = \alpha_1 + 2\alpha_2 t + \cdots + n\alpha_n t^{n-1} . \quad (14.25)
\]

Note that this definition works even over finite fields. We write \( f^{(k)} \) to mean the \( k \)-th derivative of \( f \), defined inductively as \( f^{(0)} = f \) and \( f^{(k+1)} = (f^{(k)})' \).

Proposition 14.4.69 (Linearity of differentiation). Let \( f \) and \( g \) be a polynomials and let \( \zeta \) be a scalar. Then

(a) \( (f + g)' = f' + g' \),

(b) \( (\zeta f)' = \zeta f' \).

Proposition 14.4.70 (Product rule). Let \( f \) and \( g \) be a polynomials and let \( \zeta \) be a scalar. Then

\[
 (fg)' = f'g + fg' \quad (14.26)
\]

Definition 14.4.71 (Composition of polynomials). Let \( f \) and \( g \) be polynomials. Then the composition of \( f \) with \( g \), denoted \( f \circ g \) is the polynomial obtained by replacing all occurrences of the symbol \( t \) in the expression for \( f \) with \( g \), i.e., \( f \circ g = f(g) \) (we “substitute \( g \) into \( f \)).

Proposition 14.4.72. Let \( f \) and \( g \) be polynomials and let \( \zeta \) be a scalar. Then

\[
 (f \circ g)(\zeta) = f(g(\zeta)) . \quad (14.27)
\]

Proposition 14.4.73 (Chain Rule). Let \( f, g \in \mathbb{F}[t] \) and let \( h = f \circ g \). Then

\[
 h' = (f' \circ g) \cdot g' . \quad (14.28)
\]

Proposition 14.4.74.

(a) Let \( \mathbb{F} \) be \( \mathbb{Q} \), \( \mathbb{R} \), or \( \mathbb{C} \). Let \( f \in \mathbb{F}[t] \) and let \( \zeta \in \mathbb{F} \). Then \( (t - \zeta)^k \mid f \) if and only if

\[
 f(\zeta) = f'(\zeta) = \cdots = f^{(k-1)}(\zeta) = 0
\]

(b) This is false if \( \mathbb{F} = \mathbb{F}_p \).

Proposition 14.4.75. Let \( f \in \mathbb{C}[t] \). Then \( f \) has no multiple roots if and only if \( \gcd(f, f') = 1 \).

Exercise 14.4.76. Let \( n \geq 1 \). Prove that the polynomial \( f = t^n + t + 1 \) has no multiple roots in \( \mathbb{C} \).

\footnote{This exercise holds for all subfields of \( \mathbb{C} \) and more generally for all fields of characteristic 0.}
Chapter 15

(F) Vector Spaces: Basic Concepts

15.1 Vector spaces

Many sets with which we are familiar have the desirable property that they are closed under “linear combinations” of their elements. Consider, for example, the set \( \mathbb{R}^R \) of real functions \( f : \mathbb{R} \rightarrow \mathbb{R} \). If \( f, g \in \mathbb{R}^R \) and \( \alpha, \beta \in \mathbb{R} \), then \( \alpha f + \beta g \in \mathbb{R}^R \). The same is true for the set \( C[0, 1] \) of continuous functions \( f : [0, 1] \rightarrow \mathbb{R} \) as well as many other sets, like the set \( \mathbb{R}[t] \) of polynomials with real coefficients (where \( t \) denotes the variable), the set \( \mathbb{R}^N \) of sequences of real numbers, and the 2- and 3-dimensional geometric spaces \( G_2 \) and \( G_3 \). We formalize the common properties of these sets, endowed with the notion of linear combination, with the concept of a vector space.

Let \( \mathbb{F} \) be a (finite or infinite) field. The reader not familiar with fields may think of \( \mathbb{F} \) as being \( \mathbb{R} \) or \( \mathbb{C} \).

Definition 15.1.1 (Vector space). A vector space \( V \) over a field \( \mathbb{F} \) is a set \( V \) with certain operations. We refer to the elements of \( V \) as “vectors” and the elements of \( \mathbb{F} \) as “scalars.” We assume a binary operation of addition of vectors, and an operation of multiplication of a vector by a scalar, satisfying the following axioms.

(a) \((V, +)\) is an abelian group (see Sec. 14.2).
In other words,

(b1) For all \( u, v \in V \) there exists a unique element \( u + v \in V \)

(b2) Addition is commutative: for all \( u, v \in V \) we have \( u + v = v + u \)

(b3) Addition is associative: for all \( u, v, w \in V \) we have \( u + (v + w) = (u + v) + w \)

(b4) There is a zero element, denoted \( 0 \), such that for all \( v \in V \) we have \( 0 + v = v \)

(b5) Every element \( v \in V \) has an additive inverse, denoted \( -v \), such that \( v + (-v) = 0 \).

(b) \( V \) comes with a scaling function \( V \times \mathbb{F} \rightarrow V \) such that

(b1) For all \( \alpha \in \mathbb{F} \) and \( v \in V \), there exists a unique vector \( \alpha v \in V \)

(b2) For all \( \alpha, \beta \in \mathbb{F} \) and \( v \in V \), \( (\alpha \beta)v = \alpha (\beta v) \)
(b3) For all $\alpha, \beta \in F$ and $v \in V$, $(\alpha + \beta)v = \alpha v + \beta v$

(b4) For every $\alpha \in F$ and $u, v \in V$, $\alpha(u + v) = \alpha u + \alpha v$

(b5) For every $v \in V$, $1 \cdot v = v$ (normalization)

The zero vector in a vector space $V$ is written as $0_V$, but we often just write $0$ when the context is clear.

Property (b2) is referred to as “pseudo-associativity,” because it is a form of associativity in which we are dealing with different operations (multiplication in $F$ and scaling of vectors). Similarly, Properties (b3) and (b4) are both types of “pseudo-distributivity.”

**Proposition 15.1.2.** Let $V$ be a vector space over the field $F$. For all $v \in V$, $\alpha \in F$ we have

(a) $0v = 0$.

(b) $\alpha 0 = 0$.

(c) $\alpha v = 0$ if and only if either $\alpha = 0$ or $v = 0$.

**Exercise 15.1.3.** Let $V$ be a vector space and let $x \in V$. Show that $x + x = x$ if and only if $x = 0$.

**Example 15.1.4** (Elementary geometry). The most natural examples of vector spaces are the geometric spaces $G_2$ and $G_3$. We write $G_2$ for the plane and $G_3$ for the “space” familiar from elementary geometry. We think of $G_2$ and $G_3$ as having a special point called the origin. We view the points of $G_2$ and $G_3$ as “vectors” (line segments from the origin to the point). Addition is defined by the parallelogram rule and multiplication by scalars (over $\mathbb{R}$) by scaling. Observe that $G_2$ and $G_3$ are vector spaces over $\mathbb{R}$. These classical geometries form the foundation of our intuition about vector spaces.

Note that $G_2$ is not the same as $\mathbb{R}^2$. Vectors in $G_2$ are directed segments (geometric objects), while the vectors in $\mathbb{R}^2$ are pairs of numbers. The connection between these two is one of the great discoveries of the mathematics of the modern era (Descartes).

**Examples 15.1.5.** Show that the following are vector spaces over $\mathbb{R}$.

(a) $\mathbb{R}^n$

(b) $M_n(\mathbb{R})$

(c) $\mathbb{R}^{k \times n}$

(d) $C[0, 1]$, the space of continuous real-valued functions $f : [0, 1] \to \mathbb{R}$

(e) The space $\mathbb{R}^N$ of infinite sequences of real numbers

(f) The space $\mathbb{R}^\mathbb{R}$ of real functions $f : \mathbb{R} \to \mathbb{R}$

(g) The space $\mathbb{R}[t]$ of polynomials in one variable with real coefficients.

(h) For all $k \geq 0$, the space $P_k(\mathbb{R})$ of polynomials of degree at most $k$ with coefficients in $\mathbb{R}$.
(i) The space $\mathbb{R}^\Omega$ of functions $f : \Omega \to \mathbb{R}$ where $\Omega$ is an arbitrary set.

In Section 11.1, we defined the notion of a linear form over $\mathbb{F}^n$ (Def. 11.1.1). This generalizes immediately to vector spaces over $\mathbb{F}$.

**Definition 15.1.6 (Linear form).** Let $V$ be a vector space over $\mathbb{F}$. A **linear form** is a function $f : V \to \mathbb{F}$ with the following properties.

(a) $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{F}^n$;

(b) $f(\lambda x) = \lambda f(x)$ for all $x \in \mathbb{F}^n$ and $\lambda \in \mathbb{F}$.

**Definition 15.1.7 (Dual space).** Let $V$ be a vector space over $\mathbb{F}$. The set of linear forms $f : V \to \mathbb{F}$ is called the **dual space** of $V$ and is denoted $V^*$.

**Exercise 15.1.8.** Let $V$ be a vector space over $\mathbb{F}$. Show that $V^*$ is also a vector space over $\mathbb{F}$.

In Section 1.1, we defined linear combinations of column vectors (Def. 1.1.13). This is easily generalized to linear combinations of vectors in any vector space.

**Definition 15.1.9 (Linear combination).** Let $V$ be a vector space over $\mathbb{F}$, and let $v_1, \ldots, v_k \in V$, $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$. Then $\sum_{i=1}^{k} \alpha_i v_i$ is called a **linear combination** of the vectors $v_1, \ldots, v_k$. The linear combination for which all coefficients are zero is the **trivial linear combination**.

**Exercise 15.1.10 (Empty linear combination).** What is the linear combination of the empty set? Convention 14.2.6 explains our convention for the empty sum.

**Exercise 15.1.11.**

(a) Express the polynomial $t - 1$ as a linear combination of the polynomials $t^2 - 1$, $(t - 1)^2$, $t^2 - 3t + 2$.

(b) Give an elegant proof that the polynomial $t^2 + 1$ cannot be expressed as a linear combination of the polynomials $t^2 - 1$, $(t - 1)^2$, $t^2 - 3t + 2$.

**Exercise 15.1.12.** For $\alpha \in \mathbb{R}$, express $\cos(t + \alpha)$ as a linear combination of $\cos t$ and $\sin t$.

### 15.2 Subspaces and span

In Section 1.2, we studied subspaces and span in the context of $\mathbb{F}^k$. We now generalize this to arbitrary vector spaces.

For this section, we will take $V$ to be a vector space over a field $\mathbb{F}$ and we will let $W, S \subseteq V$.

**Definition 15.2.1 (Subspace).** $W \subseteq V$ is a **subspace** (notation: $W \leq V$) if $W$ is closed under linear combinations.

**Proposition 15.2.2.** Let $W \leq V$. Then $W$ is also a vector space.

**Exercise 15.2.3.** Verify that Propositions 1.2.3 and 1.2.9 hold when $\mathbb{F}^n$ is replaced by $V$. 
Exercise 15.2.4. Describe all subspaces of the geometric spaces $G_2$ and $G_3$.

Exercise 15.2.5. Determine which of the following are subspaces of $\mathbb{R}[t]$, the space of polynomials in one variable over $\mathbb{R}$.

- (a) $\{ f \in \mathbb{R}[t] \mid \deg(f) = 5 \}$
- (b) $\{ f \in \mathbb{R}[t] \mid \deg(f) \leq 5 \}$
- (c) $\{ f \in \mathbb{R}[t] \mid f(1) = 1 \}$
- (d) $\{ f \in \mathbb{R}[t] \mid f(1) = 0 \}$
- (e) $\{ f \in \mathbb{R}[t] \mid f(\sqrt{2}) = 0 \}$
- (f) $\{ f \in \mathbb{R}[t] \mid f(\sqrt{-1}) = 0 \}$
- (g) $\{ f \in \mathbb{R}[t] \mid f(1) = f(2) \}$
- (h) $\{ f \in \mathbb{R}[t] \mid f(1) = (f(2))^2 \}$
- (i) $\{ f \in \mathbb{R}[t] \mid f(1)f(2) = 0 \}$
- (j) $\{ f \in \mathbb{R}[t] \mid f(1) = 3f(2) + 4f(3) \}$
- (k) $\{ f \in \mathbb{R}[t] \mid f(1) = 3f(2) + 4f(3) + 1 \}$
- (l) $\{ f \in \mathbb{R}[t] \mid f(1) \leq f(2) \}$

Definition 15.2.6 (Span). Let $V$ be a vector space and let $v_1, \ldots, v_m \in V$. Then the span of $S = \{ v_1, \ldots, v_m \}$, denoted $\text{span}(v_1, \ldots, v_m)$, is the smallest subspace of $V$ containing $S$, i.e.,

- (a) $\text{span} S \supseteq S$;
- (b) $\text{span} S$ is closed under linear combinations

(c) for every subspace $W \subseteq V$, if $S \subseteq W$ then $\text{span} S \subseteq W$.

Exercise 15.2.7. Repeat the exercises of Section 1.2 replacing $\mathbb{F}^k$ by $V$.

15.3 Linear independence and bases

Let $V$ be a vector space. In Section 1.3 we defined the notion of linear independence of matrices ($\mathbb{F}$ Def. 1.3.5). We now generalize this to linear independence of a list (1.3.1) of vectors in a general vector space.

Definition 15.3.1 (Linear independence). The list $(v_1, \ldots, v_k)$ of vectors in $V$ is said to be linearly independent over $\mathbb{F}$ if the only linear combination equal to $0$ is the trivial linear combination. The list $(v_1, \ldots, v_k)$ is linearly dependent if it is not linearly independent.

Definition 15.3.2. If a list $(v_1, \ldots, v_k)$ of vectors is linearly independent (dependent), we say that the vectors $v_1, \ldots, v_k$ are linearly independent (dependent).

Definition 15.3.3. We say that a set of vectors is linearly independent if a list formed by its elements (in any order and without repetitions) is linearly independent.

Exercise 15.3.4. Show that the following sets are linearly independent over $\mathbb{Q}$.

- (a) $\{ 1, \sqrt{2}, \sqrt{3} \}$
15.3. LINEAR INDEPENDENCE AND BASES

(b) \{\sqrt{x} \mid x \text{ is square-free}\} (an integer \(n\) is square-free if there is no perfect square \(k \neq 1\) such that \(k \mid n\)).

Exercise 15.3.5. Let \(U_1, U_2 \leq V\) with \(U_1 \cap U_2 = \{0\}\). Let \(v_1, \ldots, v_k \in U_1\) and \(w_1, \ldots, w_\ell \in U_2\). If the lists \((v_1, \ldots, v_k)\) and \((w_1, \ldots, w_\ell)\) are linearly independent, then so is the list \((v_1, \ldots, v_k, w_1, \ldots, w_\ell)\).

Definition 15.3.6 (Rank and dimension). The rank of a set of vectors is the maximum number of linearly independent vectors among them. For a vector space \(V\), the dimension of \(V\) is its rank, that is, \(\dim V = \text{rk} V\).

Exercise 15.3.7.

(a) When are two vectors in \(G_2\) linearly dependent?

(b) When are two vectors in \(G_3\) linearly dependent?

(c) When are three vectors in \(G_3\) linearly dependent?

Phrase your answers in geometric terms.

Definition 15.3.8. We say that the vector \(w\) depends on the list \((v_1, \ldots, v_k)\) of vectors if \(w \in \text{span}(v_1, \ldots, v_k)\), i.e., if \(w\) can be expressed as a linear combination of the \(v_i\).

Definition 15.3.9 (Linear independence of an infinite list). We say that an infinite list \((v_i \mid i \in I)\) (where \(I\) is an index set) is linearly independent if every finite sublist \((v_i \mid i \in J)\) (where \(J \subseteq I\) and \(|J| < \infty\)) is linearly independent.

Exercise 15.3.10. Verify that Exercises 13.11-13.26 hold in general vector spaces (replace \(F^n\) by \(V\) where necessary).

Example 15.3.11. For \(k = 0, 1, 2, \ldots,\) let \(f_k\) be a polynomial of degree \(k\). Show that the infinite list \((f_0, f_1, f_2, \ldots)\) is linearly independent.

Exercise 15.3.12. Find three nonzero vectors in \(G_3\) that are linearly dependent but no two are parallel.

Exercise 15.3.13. Prove that for all \(\alpha, \beta \in \mathbb{R}\), the functions \(\sin(t), \sin(t+\alpha), \sin(t+\beta)\) are linearly dependent (as members of the function space \(\mathbb{R}^\mathbb{R}\)).

Exercise 15.3.14. Let \(\alpha_1 < \alpha_2 < \cdots < \alpha_n \in \mathbb{R}\). Consider the vectors

\[
\mathbf{v}_i = \begin{pmatrix} 
\alpha_{i1} \\
\alpha_{i2} \\
\vdots \\
\alpha_{in} 
\end{pmatrix}
\] (15.1)

for \(i \geq 0\) (recall the convention that \(\alpha^0 = 1\) even if \(\alpha = 0\)). Show that \((\mathbf{v}_0, \ldots, \mathbf{v}_{n-1})\) is linearly independent.

Exercise 15.3.15 (Moment curve). Find a continuous curve in \(\mathbb{R}^n\), i.e., a continuous injective function \(f : \mathbb{R} \to \mathbb{R}^n\), such that every set of \(n\) points on the curve is linearly independent. The simplest example is called the "moment curve," and we bet you will find it.
Exercise 15.3.16. Let \( \alpha_1 < \alpha_2 < \cdots < \alpha_n \in \mathbb{R} \), and define the degree-\( n \) polynomial
\[
f = \prod_{j=1}^{n} (t - \alpha_j)
\]
For each \( 1 \leq i \leq n \), define the polynomial \( g_i \) of degree \( n - 1 \) by
\[
g_i = \frac{f}{t - \alpha_i} = \prod_{j=1, j \neq i}^{n} (t - \alpha_j).
\]
Prove that the polynomials \( g_1, \ldots, g_n \) are linearly independent.

Definition 15.3.17. We say that a set \( S \subseteq V \) generates \( V \) if \( \text{span}(S) = V \). If \( S \) generates \( V \), then \( S \) is said to be a set of generators of \( V \).

Definition 15.3.18 (Finite-dimensional vector space). We say that the vector space \( V \) is finite dimensional if \( V \) has a finite set of generators. A vector space which is not finite dimensional is infinite dimensional.

Exercise 15.3.19. Show that \( \mathbb{F}[t] \) is infinite dimensional.

Definition 15.3.20 (Basis). A list \( \mathcal{E} = (e_1, \ldots, e_k) \) of vectors is a basis of \( V \) if \( \mathcal{E} \) is linearly independent and generates \( V \).

In Section 1.3, we defined the standard basis of the space \( \mathbb{F}^n \) (Def. 1.3.34). However, general vector spaces do not have the notion of a standard basis.

Note that a list of vectors is not the same as a set of vectors, but a list of vectors which is linearly independent necessarily has no repeated elements. Note further that lists carry with them an inherent ordering; that is, bases are ordered.

Examples 15.3.21.

(a) Show that the polynomials \( 1, t, t^2, \ldots, t^k \) form a basis for \( P_k[\mathbb{F}] \).

(b) Show that the polynomials \( t^2 + t + 1, t^2 - 2t + 2, t^2 - t - 1 \) form a basis of \( P_2[\mathbb{F}] \).

(c) Express the polynomial \( f = 1 \) as a linear combination of these basis vectors.

Examples 15.3.22. For each of the following sets \( S \), describe the vectors in \( \text{span}(S) \) and give a basis for \( \text{span}(S) \).

(a) \( S = \{t, t^2\} \subseteq \mathbb{F}[t] \)

(b) \( S = \{\sin(t), \cos(t), \cos(2t), e^{it}\} \subseteq \mathbb{C}^\mathbb{R} \)

(c) \( S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ -7 \\ 4 \end{pmatrix} \right\} \subseteq \mathbb{F}^3 \)

(d) \( S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & -5 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 0 & -7 \end{pmatrix} \right\} \subseteq M_2(\mathbb{F}) \)
15.3. LINEAR INDEPENDENCE AND BASES

(e) \( S = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \right\} \subseteq M_3(\mathbb{F}) \)

Example 15.3.23. Show that the polynomials \( t^2, (t+1)^2, \) and \( (t+2)^2 \) form a basis for \( P_2[\mathbb{F}] \). Express the polynomial \( t \) in terms of this basis, and write its coordinate vector.

Exercise 15.3.24. For \( \alpha \in \mathbb{R} \), write the coordinate vector of \( \cos(t + \alpha) \) in the basis \( (\cos t, \sin t) \).

Exercise 15.3.25. Find a basis of the 0-weight subspace of \( \mathbb{R}^k \) (the 0-weight subspace is defined in \( \mathbb{R} \) Ex. 1.2.7).

Exercise 15.3.26.

(a) Find a basis of \( M_n(\mathbb{F}) \).

(b) Find a basis of \( M_n(\mathbb{F}) \) consisting of non-singular matrices.

Warning. Part [b] is easier for fields of characteristic 0 than for fields of finite characteristic.

Proposition 15.3.27. Let \( \mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_n) \) be a list of vectors in \( V \). Then \( \mathbf{b} \) is a basis of \( V \) if and only if every vector can be uniquely expressed as a linear combination of the \( \mathbf{b}_i \), i.e., for every \( \mathbf{v} \in V \), there exists a unique list of coefficients \( (\alpha_1, \ldots, \alpha_n) \) such that \( \mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{b}_i \).

Definition 15.3.28 (Maximal linearly independent set). A linearly independent set \( S \subseteq V \) is maximal if, for all \( \mathbf{v} \in V \setminus S \), \( S \cup \{ \mathbf{v} \} \) is not linearly independent.

Proposition 15.3.29. Let \( \mathbf{e} \) be a list of vectors in a vector space \( V \). Then \( \mathbf{e} \) is a basis of \( V \) if and only if it is a maximal linearly independent set.

Proposition 15.3.30. Let \( V \) be a vector space. Then \( V \) has a basis (Zorn’s lemma is needed for the infinite-dimensional case).

Proposition 15.3.31. Let \( \mathbf{e} \) be a list of vectors in \( V \). Then it is possible to extend \( \mathbf{e} \) to a basis of \( V \), that is, there exists a basis of \( V \) which has \( \mathbf{e} \) as a sublist.

Proposition 15.3.32. Let \( V \) be a vector space and let \( S \subseteq V \) be a set of generators of \( V \). Then there exists a list \( \mathbf{e} \) of vectors in \( S \) such that \( \mathbf{e} \) is a basis of \( V \).

Definition 15.3.33 (Coordinates). The coefficients \( \alpha_1, \ldots, \alpha_n \) of Ex. 15.3.27 are called the coordinates of \( \mathbf{v} \) with respect to the basis \( \mathbf{b} \).

Definition 15.3.34 (Coordinate vector). Let \( \mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_k) \) be a basis of the vector space \( V \), and let \( \mathbf{v} \in V \). Then the column vector representation of \( \mathbf{v} \) with respect to the basis \( \mathbf{b} \), or the coordinization of \( \mathbf{v} \) with respect to \( \mathbf{b} \), denoted by \( [\mathbf{v}]_\mathbf{b} \), is obtained by arranging the coordinates of \( \mathbf{v} \) with respect to \( \mathbf{b} \) in a column,
15.4 The First Miracle of Linear Algebra

In Section 1.3, we proved the First Miracle of Linear Algebra for $\mathbb{F}^n$ (Theorem 1.3.40). This generalizes immediately to abstract vector spaces.

Theorem 15.4.1 (First Miracle of Linear Algebra). Let $v_1, \ldots, v_k$ be linearly independent with $v_i \in \text{span}(w_1, \ldots, w_m)$ for all $i$. Then $k \leq m$.

The proof of this theorem requires the following lemma.

Lemma 15.4.2 (Steinitz exchange lemma). Let $(v_1, \ldots, v_k)$ be a linearly independent list such that $v_i \in \text{span}(w_1, \ldots, w_m)$ for all $i$. Then there exists $j$ ($1 \leq j \leq m$) such that the list $(w_j, v_2, \ldots, v_k)$ is linearly independent.

Corollary 15.4.3. Let $V$ be a vector space. All bases of $V$ have the same cardinality.

This is an immediate corollary to the First Miracle.

The following theorem is essentially a restatement of the First Miracle of Linear Algebra.

Theorem 15.4.4.

(a) Use the First Miracle to derive the fact that $\text{rk}(v_1, \ldots, v_k) = \dim(\text{span}(v_1, \ldots, v_k))$.

(b) Derive the First Miracle from the statement that $\text{rk}(v_1, \ldots, v_k) = \dim(\text{span}(v_1, \ldots, v_k))$.

Exercise 15.4.5. Let $V$ be a vector space of dimension $n$, and let $v_1, \ldots, v_n \in V$. The following are equivalent:

(a) $(v_1, \ldots, v_n)$ is a basis of $V$
(b) $v_1, \ldots, v_n$ are linearly independent
(c) $V = \text{span}(v_1, \ldots, v_n)$

Exercise 15.4.6. Show that $\dim(\mathbb{F}^n) = n$.

Exercise 15.4.7. Show that $\dim(\mathbb{F}^{k \times n}) = kn$.

Exercise 15.4.8. Show that $\dim(P_k) = k + 1$, where $P_k$ is the space of polynomials of degree at most $k$.

Exercise 15.4.9. What is the dimension of the subspace of $\mathbb{R}[t]$ consisting of polynomials $f$ of degree at most $n$ such that $f(\sqrt{-1}) = 0$?

Exercise 15.4.10. Show that, if $f$ is a polynomial of degree $n$, then $(f(t), f(t + 1), \ldots, f(t + n - 1))$ is a basis of $P_n[\mathbb{F}]$. 

i.e.,

$$[v]_b = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}$$ (15.2)

where $v = \sum_{i=1}^{k} \alpha_i b_i$. 

$$\text{rk}(v_1, \ldots, v_k) = \dim(\text{span}(v_1, \ldots, v_k))$$

Exercise 15.4.11. Show that any list of polynomials, one of each degree 0, ..., n, forms a basis of $P_n[\mathbb{F}]$.

Proposition 15.4.12. Let $V$ be an $n$-dimensional vector space with subspaces $U_1, U_2$ such that $U_1 \cap U_2 = \{0\}$. Then

$$\dim U_1 + \dim U_2 \leq n. \quad (15.3)$$

Proposition 15.4.13 (Modular equation). Let $V$ be a vector space, and let $U_1, U_2 \leq V$. Then

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2. \quad (15.4)$$

Exercise 15.4.14. Let $A = (\alpha_{ij}) \in \mathbb{R}^{n \times n}$, and assume the columns of $A$ are linearly independent. Prove that it is always possible to change the value of an entry in the first row so that the columns of $A$ become linearly dependent.

Exercise 15.4.15. Call a sequence $(a_0, a_1, a_2, \ldots)$ “Fibonacci-like” if for all $n$, $a_{n+2} = a_{n+1} + a_n$.

(a) Prove that Fibonacci-like sequences form a 2-dimensional vector space.

(b) Find a basis for the space of Fibonacci-like sequences.

(c) Express the Fibonacci sequence $(0, 1, 1, 2, 5, 8, \ldots)$ as a linear combination of these basis vectors.

\(\heartsuit\) Exercise 15.4.16. Let $f$ be a polynomial. Prove that $f$ has a multiple $g = f \cdot h \neq 0$ in which every exponent is prime, i.e.,

$$g = \sum_{p \text{ prime}} \alpha_p x^p \quad (15.5)$$

for some coefficients $\alpha_p$.

Definition 15.4.17 (Elementary operations). The three types of elementary operations defined as “elementary column operations” in Sec. 3.2 can be applied to any list of vectors.

Proposition 15.4.18. Performing elementary operations does not change the rank of a set of vectors.

15.5 Direct sums

Definition 15.5.1 (Disjoint subspaces). Two subspaces $U_1, U_2 \leq V$ are disjoint if $U_1 \cap U_2 = \{0\}$.

Definition 15.5.2 (Direct sum). Let $V$ be a vector space and let $U_1, U_2 \leq V$. We say that $V$ is the direct sum of $U_1$ and $U_2$ (notation: $V = U_1 \oplus U_2$) if $V = U_1 + U_2$ and $U_1 \cap U_2 = \{0\}$.

More generally, we say that $V$ is the direct sum of subspaces $U_1, \ldots, U_k \leq V$ (notation: $V = U_1 \oplus \cdots \oplus U_k$) if $V = U_1 + \cdots + U_k$ and for all $i$,

$$U_i \cap \left( \sum_{j \neq i} U_j \right) = \{0\}. \quad (15.6)$$
Proposition 15.5.3. Let $V$ be a vector space and let $U_1, \ldots, U_k \leq V$. Then

$$\dim(U_1 \oplus \cdots \oplus U_k) = \sum_{i=1}^{k} \dim U_i . \quad (15.7)$$

The following exercise shows that this could actually be used as the definition of the direct sum in finite-dimensional spaces, but that statement in fact holds in infinite-dimensional spaces as well.

Proposition 15.5.4. Let $V$ be a finite-dimensional vector space, and let $U_1, \ldots, U_k \leq V$, with

$$\dim \left( \sum_{i=1}^{k} U_i \right) = \sum_{i=1}^{k} \dim U_i .$$

Then

$$\sum_{i=1}^{k} U_i = \bigoplus_{i=1}^{k} U_i .$$

Proposition 15.5.5. Let $V$ be a vector space and let $U_1, \ldots, U_k \leq V$. Then $W = \sum_{i=1}^{k} U_i$ is a direct sum if and only if for every choice of $k$ vectors $u_i \ (i = 1, \ldots, k)$ where $u_i \in U_i \setminus \{0\}$, the vectors $u_1, \ldots, u_k$ are linearly independent.

We note that the notion of direct sum extends the notion of linear independence to subspaces.

Proposition 15.5.6. The vectors $v_1, \ldots, v_k$ are linearly independent if and only if

$$\sum_{i=1}^{k} \text{span}(v_i) = \bigoplus_{i=1}^{k} \text{span}(v_i) . \quad (15.8)$$
Chapter 16

(\mathbb{F}) Linear Maps

16.1 Linear map basics

Definition 16.1.1 (Linear map). Let \( V \) and \( W \) be vector spaces over the same field \( \mathbb{F} \). A function \( \varphi : V \to W \) is called a linear map or homomorphism if for all \( v, w \in V \) and \( \alpha \in \mathbb{F} \)

(a) \( \varphi(\alpha v) = \alpha \varphi(v) \)

(b) \( \varphi(v + w) = \varphi(v) + \varphi(w) \)

Exercise 16.1.2. Show that if \( \varphi : V \to W \) is a linear map, then \( \varphi(0_V) = 0_W \).

Proposition 16.1.3. Linear maps preserve linear combinations, i.e.,

\[
\varphi \left( \sum_{i=1}^{k} \alpha_i v_i \right) = \sum_{i=1}^{k} \alpha_i \varphi(v_i) \quad (16.1)
\]

Exercise 16.1.4. True or false?

(a) If the vectors \( v_1, \ldots, v_k \) are linearly independent, then \( \varphi(v_1), \ldots, \varphi(v_k) \) are linearly independent.

(b) If the vectors \( v_1, \ldots, v_k \) are linearly dependent, then \( \varphi(v_1), \ldots, \varphi(v_k) \) are linearly dependent.

Example 16.1.5. Our prime examples of linear maps are those defined by matrix multiplication. Every matrix \( A \in \mathbb{F}^{k \times n} \) defines a linear map \( \varphi_A : \mathbb{F}^n \to \mathbb{F}^k \) by \( x \mapsto Ax \).

Example 16.1.6. The projection of \( \mathbb{R}^3 \) onto \( \mathbb{R}^2 \) defined by

\[
\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \mapsto \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (16.2)
\]

is a linear map (verify!).

Example 16.1.7. The map \( f \mapsto \int_0^1 f(t)dt \) is a linear map from \( C[0, 1] \to \mathbb{R} \) (verify!).

Example 16.1.8. Differentiation \( \frac{d}{dt} \) is a linear map \( \frac{d}{dt} : P_n[\mathbb{F}] \to P_{n-1}[\mathbb{F}] \) from the space of polynomials of degree \( \leq n \) to the space of polynomials of degree \( \leq n - 1 \) (verify!).

Example 16.1.9. The map \( \varphi : P_n[\mathbb{R}] \to \mathbb{R}^k \) defined by

\[
f \mapsto \begin{pmatrix} f(\alpha_1) \\ f(\alpha_2) \\ \vdots \\ f(\alpha_k) \end{pmatrix} \quad (16.3)
\]

where \( \alpha_1 < \alpha_2 < \cdots < \alpha_n \) is a linear map (verify!).
Example 16.1.10. Let us fix $n + 1$ distinct real numbers $\alpha_0 < \alpha_1 < \cdots < \alpha_n$. Let $f \in \mathbb{R}$ be a real function. Interpolation is the map $f \mapsto L(f) \in P_n(\mathbb{R})$ where $L(f) = p$ is the unique polynomial of degree $\leq n$ with the property that $p(\alpha_i) = f(\alpha_i)$ $(i = 0, \ldots, n)$. This is a linear map from $\mathbb{R} \to P_n(\mathbb{R})$ (verify!).

Example 16.1.11. The map $\varphi: P_n(\mathbb{R}) \to T_n = \text{span}\{1, \cos w, \cos 2w, \ldots, \cos nw\}$ defined by
\[
\varphi(\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^n) = \alpha_0 + \alpha_1 \cos w + \alpha_2 \cos 2w + \cdots + \alpha_n \cos n t
\]
(16.4)
is a linear map from polynomials to trigonometric polynomials (verify!).

Notation 16.1.12. The set of linear maps $\varphi: V \to W$ is denoted by $\text{Hom}(V,W)$.

Fact 16.1.13. $\text{Hom}(V,W)$ is a subspace of the function space $W^V$.

Definition 16.1.14 (Composition of linear maps). Let $U$, $V$, and $W$ be vector spaces, and let $\varphi : U \to V$ and $\psi : V \to W$ be linear maps. Then the composition of $\psi$ with $\varphi$, denoted by $\psi \circ \varphi$ or $\varphi \varphi$, is the map $\eta : U \to W$ defined by
\[
\eta(v) := \psi(\varphi(v))
\]
(16.5)

FIGURE: $U \xrightarrow{\varphi} V \xrightarrow{\psi} W$

Proposition 16.1.15. Let $U$, $V$, and $W$ be vector spaces and let $\varphi : U \to V$ and $\psi : V \to W$ be linear maps. Then $\varphi \circ \psi$ is a linear map.

Linear maps are uniquely determined by their action on a basis, and we are free to choose this action arbitrarily. This is more formally expressed in the following theorem.

Theorem 16.1.16 (Degree of freedom of linear maps). Let $V$ and $W$ be vector spaces with $e = (e_1, \ldots, e_k)$ a basis of $V$, and $w_1, \ldots, w_k$ arbitrary vectors in $W$. Then there exists a unique linear map $\varphi : V \to W$ such that $\varphi(v_i) = w_i$ for $1 \leq i \leq k$.

Exercise 16.1.17. Show that $\dim(\text{Hom}(V,W)) = \dim(V) \cdot \dim(W)$.

In Section 15.3 we represented vectors by the column vectors of their coordinates with respect to a given basis. As the next step in translating geometric objects to tables of numbers, we assign matrices to linear maps relative to given bases in the domain and the target space. Our key tool for this endeavor is Theorem 16.1.16.

16.2 Isomorphisms

Let $V$ and $W$ be vector spaces over the same field.

Definition 16.2.1 (Isomorphism). A linear map $\varphi \in \text{Hom}(V,W)$ is said to be an isomorphism if it is a bijection. If there exists an isomorphism between $V$ and $W$, then $V$ and $W$ are said to be isomorphic. The circumstance that “$V$ is isomorphic to $W$” is denoted $V \cong W$. 
16.3. THE RANK-NULLITY THEOREM

**Fact 16.2.2.** The inverse of an isomorphism is an isomorphism.

**Fact 16.2.3.** Isomorphisms preserve linear independence and, moreover, map bases to bases.

**Fact 16.2.4.** Let \( V \) and \( W \) be vector spaces, let \( \varphi : V \to W \) be an isomorphism, and let \( v_1, \ldots, v_k \in V \). Then
\[
\text{rk}(v_1, \ldots, v_k) = \text{rk}(\varphi(v_1), \ldots, \varphi(v_k))
\]

**Exercise 16.2.5.** Show that \( \sim \) is an equivalence relation, that is, for vector spaces \( U, V, \) and \( W \):

(a) \( V \cong V \) (reflexive)

(b) If \( V \cong W \) then \( W \cong V \) (symmetric)

(c) If \( U \cong V \) and \( V \cong W \) then \( U \cong W \) (transitive)

**Exercise 16.2.6.** Let \( V \) be an \( n \)-dimensional vector space over \( \mathbb{F} \). Show that \( V \cong \mathbb{F}^n \).

**Proposition 16.2.7.** Two vector spaces over the same field are isomorphic if and only if they have the same dimension.

### 16.3 The Rank-Nullity Theorem

**Definition 16.3.1 (Image).** The image of a linear map \( \varphi \), denoted \( \text{im}(\varphi) \), is
\[
\text{im}(\varphi) := \{ \varphi(v) \mid v \in V \}
\]  

**Definition 16.3.2 (Kernel).** The kernel of a linear map \( \varphi \), denoted \( \text{ker}(\varphi) \), is
\[
\text{ker}(\varphi) := \{ v \in V \mid \varphi(v) = 0_W \} = \varphi^{-1}(0_W) \tag{16.7}
\]

**Proposition 16.3.3.** Let \( \varphi : V \to W \) be a linear map. Then \( \text{im}(\varphi) \leq W \) and \( \text{ker}(\varphi) \leq V \).

**Definition 16.3.4 (Rank of a linear map).** The rank of a linear map \( \varphi \) is defined as
\[
\text{rk}(\varphi) := \dim(\text{im}(\varphi))
\]

**Definition 16.3.5 (Nullity).** The nullity of a linear map \( \varphi \) is defined as
\[
\text{nullity}(\varphi) := \dim(\text{ker}(\varphi))
\]

**Exercise 16.3.6.** Let \( \varphi : V \to W \) be a linear map.

(a) Show that \( \dim(\text{im} \varphi) \leq \dim V \).

(b) Use this to reprove \( \text{rk}(AB) \leq \min\{\text{rk}(A), \text{rk}(B)\} \).

**Theorem 16.3.7 (Rank-Nullity Theorem).** Let \( \varphi : V \to W \) be a linear map. Then
\[
\text{rk}(\varphi) + \text{nullity}(\varphi) = \dim(V) \tag{16.8}
\]

**Exercise 16.3.8.** Let \( n = k + \ell \). Find a linear map \( \varphi : \mathbb{R}^n \to \mathbb{R}^n \) which has rank \( k \), and therefore has nullity \( \ell \).

**Examples 16.3.9.** Find the rank and nullity of each of the linear maps in Examples \[16.1.6\]
16.4 Linear transformations

Let $V$ be a vector space over the field $\mathbb{F}$.

In Section 16.1, we defined the notion of a linear map. We now restrict ourselves to those maps which map a vector space into itself.

**Definition 16.4.1 (Linear transformation).** If $\varphi : V \to V$ is a linear map, then $\varphi$ is said to be a linear transformation or a linear operator.

**Definition 16.4.2 (Identity transformation).** The identity transformation of $V$ is the transformation $\text{id} : V \to V$ defined by $\text{id}(v) = v$ for all $v \in V$.

**Definition 16.4.3 (Scalar transformation).** The scalar transformations of $V$ are the transformations of the form $v \mapsto \alpha v$ for a fixed $\alpha \in \mathbb{F}$. In other words, a scalar transformation can be written as $\alpha \cdot \text{id}$.

**Examples 16.4.4.** The following are linear transformations of the 2-dimensional geometric space $G_2$ (verify!).

(a) Rotation about the origin by an angle $\theta$

(b) Reflection about any line passing through the origin

(c) Shearing parallel to a line (tilting the deck)

(d) Scaling by $\alpha$ in one direction

**Examples 16.4.5.** The following are linear transformations of the 3-dimensional geometric space $G_3$ (verify!).

(a) Rotation by $\theta$ about any line

(b) Projection into any plane

(c) Reflection through any plane passing through the origin

(d) Central reflection

**Example 16.4.6.** Let $S_\alpha : \mathbb{R}^\mathbb{R}$ be the left shift by $\alpha$ operator, defined by $S_\alpha(f) = g$ where

$$g(t) := f(t + \alpha) . \quad (16.9)$$

$S_\alpha$ is a linear transformation (verify!).

**Examples 16.4.7.** The following are linear transformations of $P_n[\mathbb{F}]$ (verify!).

(a) Differentiation $\frac{d}{dt}$

(b) Left shift by $\alpha$, $f \mapsto S_\alpha(f)$

(c) Multiplication by $t$, followed by differentiation, i.e., the map $f \mapsto (tf)'$

**Examples 16.4.8.** The following are linear transformations of the space $\mathbb{R}^N$ of infinite sequences of real numbers (verify!).

(a) Left shift,

$$(\alpha_0, \alpha_1, \alpha_2, \ldots) \mapsto (\alpha_1, \alpha_2, \alpha_3, \ldots)$$

(b) Difference, $(\alpha_0, \alpha_1, \alpha_2, \ldots) \mapsto (\alpha_1 - \alpha_0, \alpha_2 - \alpha_1, \alpha_3 - \alpha_2, \ldots)$
Example 16.4.9. Let $V = \text{span}(\sin t, \cos t)$. Differentiation is a linear transformation of $V$ (verify!).

Exercise 16.4.10. The difference operator $\Delta : P_k[\mathbb{F}] \to P_k[\mathbb{F}]$, defined by

$$\Delta f := S_\alpha(f) - f$$

is linear, and $\text{deg}(\Delta f) = \text{deg}(f) - 1$ if $\text{deg}(f) \geq 1$.

16.4.1 Eigenvectors, eigenvalues, eigenspaces

In Chapter 8, we discussed the notion of eigenvectors and eigenvalues of square matrices. This is easily generalized to eigenvectors and eigenvalues of linear transformations.

Definition 16.4.11 (Eigenvector). Let $\varphi : V \to V$ be a linear transformation. Then $v \in V$ is an eigenvector of $\varphi$ if $v \neq 0$ and there exists $\lambda \in \mathbb{F}$ such that $\varphi(v) = \lambda v$. In this case we say that $v$ is an eigenvector to eigenvalue $\lambda$.

Definition 16.4.12 (Eigenvalue). Let $\varphi : V \to V$ be a linear transformation. Then $\lambda \in \mathbb{F}$ is an eigenvalue of $\varphi$ if there exists a nonzero vector $v \in V$ such that $\varphi(v) = \lambda v$.

Compare these definitions with those in Chapter 8. In particular, if $A$ is a square matrix, verify that its eigenvectors and eigenvalues are the same as those of the linear transformation $\varphi_A$ associated with $A$ (Example 16.1.5), i.e., the map $x \mapsto Ax$.

Exercise 16.4.13. Let $\varphi : V \to V$ be a linear transformation and let $v_1$ and $v_2$ be eigenvectors to distinct eigenvalues. Then $v_1 + v_2$ is not an eigenvector.

Exercise 16.4.14. Let $\varphi : V \to V$ be a linear transformation such that every nonzero vector is an eigenvector. Then $\varphi$ is a scalar transformation.

Exercise 16.4.15. Let $\varphi : V \to V$ be a linear transformation and let $v_1, \ldots, v_k$ be eigenvectors to distinct eigenvalues. Then the $v_i$ are linearly independent.

Definition 16.4.16 (Eigenspace). Let $V$ be a vector space and let $\varphi : V \to V$ be a linear transformation. We denote by $U_\lambda$ the set

$$U_\lambda := \{v \in V \mid \varphi(v) = \lambda v\}.$$  \hspace{1cm} (16.10)

This set is called the eigenspace corresponding to the eigenvalue $\lambda$.

Exercise 16.4.17. Let $\varphi : V \to V$ be a linear transformation. Show that, for all $\lambda \in \mathbb{F}$, $U_\lambda$ is a subspace of $V$.

Proposition 16.4.18. Let $\varphi : V \to V$ be a linear transformation. Then

$$\sum_\lambda U_\lambda = \bigoplus_\lambda U_\lambda.$$ \hspace{1cm} (16.11)

where $\oplus$ represents the direct sum (Def. 15.5.2) and the summation is over all eigenvalues.

Examples 16.4.19. Determine the rank, nullity, eigenvalues (and their geometric multiplicities), and eigenvectors of each of the transformations in Examples 16.4.4 16.4.10.
**Definition 16.4.20 (Eigenbasis).** Let \( \varphi : V \to V \) be a linear transformation. An **eigenbasis** of \( \varphi \) is a basis of \( V \) consisting of eigenvectors of \( \varphi \).

**Exercise 16.4.21.**

**16.4.2 Invariant subspaces**

**Definition 16.4.22 (Invariant subspace).** Let \( \varphi : V \to V \) and let \( W \leq V \). Then \( W \) is a **\( \varphi \)-invariant subspace** of \( V \) if, for all \( w \in W \), we have \( \varphi(w) \in W \). For any linear transformation \( \varphi \), the **trivial invariant subspace** is the subspace \( \{0\} \).

**Exercise 16.4.23.** Let \( \varphi : G_3 \to G_3 \) be a rotation about the vertical axis through the origin. What are the \( \varphi \)-invariant subspaces?

**Exercise 16.4.24.** Let \( \pi : G_3 \to G_3 \) be the projection onto the horizontal plane. What are the \( \pi \)-invariant subspaces?

**Exercise 16.4.25.**

(a) What are the invariant subspaces of id?

(b) What are the invariant subspaces of the 0 transformation?

**Exercise 16.4.26.** Let \( \varphi : V \to V \) be a linear transformation and let \( \lambda \) be an eigenvalue of \( \varphi \). What are the invariant subspaces of \( \varphi + \lambda \text{id} \)?

**Exercise 16.4.27.** Over every field \( \mathbb{F} \), find an infinite-dimensional vector space \( V \) and linear transformation \( \varphi : V \to V \) that has no finite-dimensional invariant subspaces other than \( \{0\} \).

**Proposition 16.4.28.** Let \( \varphi : V \to V \) be a linear transformation. Then \( \ker \varphi \) and \( \im \varphi \) are \( \varphi \)-invariant subspaces.

**Proposition 16.4.29.** Let \( \varphi : V \to V \) be a linear transformations and let \( W_1, W_2 \leq V \) be \( \varphi \)-invariant subspaces. Then

(a) \( W_1 + W_2 \) is a \( \varphi \)-invariant subspace;

(b) \( W_1 \cap W_2 \) is a \( \varphi \)-invariant subspace.

**Definition 16.4.30 (Restriction to a subspace).** Let \( \varphi : V \to V \) be a linear map and let \( W \leq V \) be a \( \varphi \)-invariant subspace of \( V \). The **restriction** of \( \varphi \) to \( W \), denoted \( \varphi_W \), is the linear map \( \varphi_W : W \to W \) defined by \( \varphi_W(w) = \varphi(w) \) for all \( w \in W \).

**Proposition 16.4.31.** Let \( \varphi : V \to V \) be a linear transformation and let \( W \leq V \) be a \( \varphi \)-invariant subspace. Let \( f \in \mathbb{F}[t] \). Then

\[
f(\varphi_W) = f(\varphi)_W. \quad (16.12)
\]

**Proposition 16.4.32.** Let \( \varphi : V \to V \) be a linear transformation. The following are equivalent.

(a) \( \varphi \) is a scalar transformation;

(b) all subspaces of \( V \) are \( \varphi \)-invariant;

(c) all 1-dimensional subspaces of \( V \) are \( \varphi \)-invariant;

(d) all hyperplanes (Def. 5.2.1) are \( \varphi \)-invariant.
Exercise 16.4.33. Let $S$ be the shift operator (Example 16.4.8 (a)) on the space $\mathbb{R}^N$ of sequences of real numbers, defined by

$$S(\alpha_0, \alpha_1, \alpha_2, \ldots) := (\alpha_1, \alpha_2, \alpha_3, \ldots). \quad (16.13)$$

(a) In Ex. 15.4.15, we defined the space of Fibonacci-like sequences. Show that this is an $S$-invariant subspace of $\mathbb{R}^N$.

(b) Find an eigenbasis of $S$ in this subspace.

(c) Use the result of part (b) to find an explicit formula for the $n$-th Fibonacci number.

Definition 16.4.34 (Minimal invariant subspace). Let $\varphi: V \to V$ be a linear transformation. Then $U \leq V$ is a minimal invariant subspace of $\varphi$ if the only invariant subspace properly contained in $U$ is $\{0\}$.

Exercise 16.4.35. Let $\rho = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$ be an $n \times n$ matrix.

(a) Count the invariant subspaces of $\rho$ over $\mathbb{C}$

(b) Find a 2-dimensional minimal invariant subspace of $\rho$ over $\mathbb{R}$

(c) Count the invariant subspaces of $\rho$ over $\mathbb{Q}$ if $n$ is prime

Definition 16.4.36. Let $\varphi: V \to V$ be a linear transformation. Then the $n$-th power of $\varphi$ is defined as

$$\varphi^n := \varphi \circ \cdots \circ \varphi \quad \text{(n times)} \quad (16.14)$$

Definition 16.4.37. Let $\varphi: V \to V$ be a linear transformation and let $f \in \mathbb{F}[t]$, say

$$f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n.$$ 

Then we define $f(\varphi)$ by

$$f(\varphi) := \alpha_0 \text{id} + \alpha_1 \varphi + \cdots + \alpha_n \varphi^n. \quad (16.15)$$

In particular,

$$f(\varphi)(v) = \alpha_0 v + \alpha_1 \varphi(v) + \cdots + \alpha_n \varphi^n(v). \quad (16.16)$$

Exercise 16.4.38. Let $\varphi: V \to V$ be a linear transformation and let $f \in \mathbb{F}[t]$. Verify that $f(\varphi)$ is a linear transformation of $V$.

Definition 16.4.39 (Chain of invariant subspaces). Let $\varphi: V \to V$ be a linear transformation, and let $U_1, \ldots, U_k \leq V$ be $\varphi$-invariant subspaces. We say that the $U_i$ form a chain if whenever $i \neq j$, either $U_i \leq U_j$ or $U_j \leq U_i$.

Definition 16.4.40 (Maximal chain). Let $\varphi: V \to V$ be a linear transformation and let $U_1, \ldots, U_k \leq V$ be a chain of $\varphi$-invariant subspaces. We say that this chain is maximal if, for all $W \leq V$ ($W \neq U_i$ for all $i$), the $U_i$ together with $W$ do not form a chain.
Exercise 16.4.41. Let \( \frac{d}{dt} : P_n(\mathbb{R}) \to P_n(\mathbb{R}) \) be the derivative linear transformation (Def. 14.4.68) of the space of real polynomials of degree at most \( n \).

(a) Prove that the number of \( \frac{d}{dt} \)-invariant subspaces is \( n + 2 \);

(b) give a very simple description of each;

(c) prove that they form a maximal chain.

Exercise 16.4.42. Let \( V = \mathbb{F}_p \mathbb{F}_p \), and let \( \varphi \) be the shift-by-1 operator, i.e., \( \varphi(f)(t) = f(t + 1) \). What are the invariant subspaces of \( \varphi \)? Prove that they form a maximal chain.

Proposition 16.4.43. Let \( \varphi : V \to V \) be a linear transformation, and let \( f \in \mathbb{F}[t] \). Then \( \ker f(\varphi) \) and \( \text{im } f(\varphi) \) are invariant subspaces.

The next exercise shows that invariant subspaces generalize the notion of eigenvectors.

Exercise 16.4.44. Let \( v \in V \) be a nonzero vector and let \( \varphi : V \to V \) be a linear transformation. Then \( v \) is an eigenvector of \( \varphi \) if and only if \( \text{span}(v) \) is \( \varphi \)-invariant.

Proposition 16.4.45. Let \( V \) be a vector space over \( \mathbb{R} \) and let \( \varphi : V \to V \) be a linear transformation. Then \( \varphi \) has an invariant subspace of dimension at most 2.

Proposition 16.4.46. Let \( V \) be an \( n \)-dimensional vector space and let \( \varphi : V \to V \) be a linear transformation. The following are equivalent.

(a) There exists a maximal chain of subspaces, all of which are invariant.

(b) There is a basis \( \mathbf{b} \) of \( V \) such that \( [\varphi]_{\mathbf{b}} \) is triangular.

Proposition 16.4.47. Let \( V \) be a finite-dimensional vector space with basis \( \mathbf{b} \) and let \( \varphi : V \to V \) be a linear transformation.

(a) Let \( \mathbf{b} \) a basis of \( V \). Then \( [\varphi]_{\mathbf{b}} \) is triangular if and only if every initial segment of \( \mathbf{b} \) spans a \( \varphi \)-invariant subspace of \( V \).

(b) If such a basis exists, then such an orthonormal basis exists.

Proposition 16.4.48. Let \( V \) be a finite-dimensional vector space with basis \( \mathbf{b} \) and let \( \varphi : V \to V \) be a linear transformation. Then \( [\varphi]_{\mathbf{b}} \) is diagonal if and only if \( \mathbf{b} \) is an eigenbasis of \( \varphi \).

Exercise 16.4.49. Infer Schur’s Theorem (Thm. 12.4.9) and the real version of Schur’s Theorem (Thm. 12.4.18) from the preceding exercise.

16.5 Coordinatization

In Section 15.3, we defined coordinatization of a vector with respect to some basis (Def. 15.3.34). We now extend this to the notion of coordinatization of linear maps.
Definition 16.5.1 (Coordinatization). Let $V$ be an $n$-dimensional vector space with basis $\mathcal{E} = (e_1, \ldots, e_n)$, and let $W$ be an $m$-dimensional vector space with basis $\mathcal{F} = (f_1, \ldots, f_m)$. Let $\alpha_{ij} \ (1 \leq i \leq m, 1 \leq j \leq n)$ be coefficients such that $\varphi(e_j) = \sum_{i=1}^{m} \alpha_{ij} f_i$. Then the matrix representation or coordinatization of $\varphi$ with respect to the bases $\mathcal{E}$ and $\mathcal{F}$ is the $m \times n$ matrix

$$\left[ \varphi \right]_{\mathcal{E},\mathcal{F}} := \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \quad (16.17)$$

If $\varphi : V \to V$ is a linear transformation, then we write $[\varphi]_\mathcal{E}$ instead of $[\varphi]_{\mathcal{E},\mathcal{F}}$.

So the $j$-th column of $[\varphi]_{\mathcal{E},\mathcal{F}}$ is $[\varphi(e_j)]_\mathcal{F}$, the coordinate vector of the image of the $j$-th basis vector of $V$ in the basis $\mathcal{F}$ of $W$.

Exercise 16.5.2. Write the matrix representation of each of the linear transformations in Ex. 16.4.4 in a basis consisting of perpendicular unit vectors.

Exercise 16.5.3. Compute the matrix of the “rotation by $\theta$” transformation with respect to the basis of two unit vectors at an angle $\theta$.

Exercise 16.5.4. In the preceding two exercises, we computed two matrices representing the “rotation by $\theta$” transformation, corresponding to the two different bases considered. Compare (a) the traces and (b) the determinants (Prop. 6.2.3(a)) of these two matrices.

Example 16.5.5. Write the matrix representation of each of the linear transformations in Ex. 16.4.5 in the basis consisting of three mutually perpendicular unit vectors.

Example 16.5.6. Write the matrix representation of each of the linear transformations in Ex. 16.4.7 in the basis $(1, t, t^2, \ldots, t^n)$ of the polynomial space $P_n(\mathbb{F})$.

The next exercise demonstrates that under our rules of coordinatization, the action of a linear map corresponds to multiplying a column vector by a matrix.

Proposition 16.5.7. For any $v \in V$, if $\varphi : V \to W$ is a linear map, then $[\varphi(v)]_\mathcal{F} = [\varphi]_{\mathcal{E},\mathcal{F}}[v]_\mathcal{E}$.

Moreover, coordinatization treats eigenvectors the way we would expect.

Proposition 16.5.8. Let $V$ be a vector space with basis $\mathcal{B}$ and let $\varphi : V \to V$ be a linear transformation. Write $A = [\varphi]_\mathcal{B}$. Then

(a) $A$ and $\varphi$ have the same eigenvalues

(b) $v \in V$ is an eigenvector of $\varphi$ with eigenvalue $\lambda$ if and only if $[v]_\mathcal{B}$ is an eigenvector of $A$ with eigenvalue $\lambda$.

The next exercise shows that under our coordinatization, composition of linear maps (Def. 16.1.14) corresponds to matrix multiplication. This gives a natural explanation of why we multiply matrices the way we do, and why this operation is associative.
Proposition 16.5.9. Let $U$, $V$, and $W$ be vector spaces with bases $\mathbf{e}$, $\mathbf{f}$, and $\mathbf{g}$, respectively, and let $\varphi : U \to V$ and $\psi : V \to W$ be linear maps. Then
\begin{equation}
[\psi \varphi]_{\mathbf{g} \mathbf{e}} = [\psi]_{\mathbf{f} \mathbf{g}} [\varphi]_{\mathbf{e} \mathbf{f}} \tag{16.18}
\end{equation}

Exercise 16.5.10. Explain the comment before the preceding exercise.

Exercise 16.5.11. Let $V$ and $W$ be vector spaces, with $\dim V = n$ and $\dim W = k$. Infer from Prop. 16.5.9 that coordinatization is an isomorphism between $\text{Hom}(V, W)$ and $\mathbb{F}^{k \times n}$.

Exercise 16.5.12. Let $\rho_\theta$ be the “rotation by $\theta$” linear map in $G_2$.

(a) Show that, for $\alpha, \beta \in \mathbb{R}$, $\rho_{\alpha + \beta} = \rho_\alpha \circ \rho_\beta$.

(b) Use matrix multiplication to derive the addition formulas for $\sin$ and $\cos$.

Example 16.5.13. Let $V$ be an $n$-dimensional vector space with basis $\mathbf{b}$, and let $A \in \mathbb{F}^{k \times n}$. Define $\varphi : V \to \mathbb{F}^k$ by $x \mapsto A[x]_\mathbf{b}$.

(a) Show that $\varphi$ is a linear map.

(b) Show that $\operatorname{rk} \varphi_A = \operatorname{rk} A$

Definition 16.5.14. Let $V$ be a finite-dimensional vector space over $\mathbb{R}$ and let $\mathbf{b}$ be a basis of $V$. Let $\varphi : V \to V$ be a nonsingular linear transformation. We say that $\varphi$ is sense-preserving if $\det[\varphi]_\mathbf{b} > 0$ and $\varphi$ is sense-reversing if $\det[\varphi]_\mathbf{b} < 0$.

Proposition 16.5.15.

(a) The sense-preserving linear transformations of the plane are rotations.

(b) The sense-reversing transformations of the plane are reflections about a line through the origin.

Exercise 16.5.16. What linear transformations of $G_2$ fix the origin?

Definition 16.5.17 (Rotational reflection). A rotational reflection of $G_3$ is a rotation followed by a central reflection.

Proposition 16.5.18.

(a) The sense-preserving linear transformations of 3-dimensional space are rotations about an axis.

(b) The sense-reversing linear transformations of 3-dimensional space are rotational reflections.

16.6 Change of basis

We now have the equipment necessary to discuss change of basis transformations and the result of change of basis on the matrix representation of linear maps.

Definition 16.6.1 (Change of basis transformation). Let $V$ be a vector space with bases $\mathbf{e} = (e_1, \ldots, e_n)$ and $\mathbf{e}' = (e'_1, \ldots, e'_n)$. Then the change of basis transformation (from $\mathbf{e}$ to $\mathbf{e}'$) is the linear transformation $\sigma : V \to V$ given by $\sigma(e_i) = e'_i$, for $1 \leq i \leq n$. 
Proposition 16.6.2. Let $V$ be a vector space with bases $\mathcal{e}$ and $\mathcal{e}'$, and let $\sigma : V \to V$ be the change of basis transformation from $\mathcal{e}$ to $\mathcal{e}'$. Then $[\sigma]_{\mathcal{e}} = [\sigma]_{\mathcal{e}'}$.

For this reason, we often denote the matrix representation of the change of basis transformation $\sigma$ by $[\sigma]$ rather than by, e.g., $[\sigma]_{\mathcal{e}}$.

Fact 16.6.3. Let $\sigma : V \to V$ be a change of basis transformation. Then $[\sigma]$ is invertible.

Notation 16.6.4. Let $V$ be a vector space with bases $\mathcal{e}$ and $\mathcal{e}'$. When changing basis from $\mathcal{e}$ to $\mathcal{e}'$, we sometimes refer to $\mathcal{e}$ as the “old” basis and $\mathcal{e}'$ as the “new” basis. So if $v \in V$ is a vector, we often write $[v]_{\text{old}}$ in place of $[v]_{\mathcal{e}}$ and $[v]_{\text{new}}$ in place of $[v]_{\mathcal{e}'}$. Likewise, if $W$ is a vector space with bases $\mathcal{f}$ and $\mathcal{f}'$ and we change bases from $\mathcal{f}$ to $\mathcal{f}'$, we consider $\mathcal{f}$ the “old” basis and $\mathcal{f}'$ the “new” basis. So if $v \in W$ is a vector, we write $[v]_{\text{old}}$ in place of $[v]_{\mathcal{f}}$ and $[v]_{\text{new}}$ in place of $[v]_{\mathcal{f}'}$.

Proposition 16.6.5. Let $v \in V$ and let $\mathcal{e}$ and $\mathcal{e}'$ be bases of $V$. Let $\sigma$ be the change of basis transformation from $\mathcal{e}$ to $\mathcal{e}'$. Then

$$[v]_{\text{new}} = [\sigma]^{-1}[v]_{\text{old}}.$$ \hfill (16.19)

Numerical exercise 16.6.6. For each of the following vector spaces $V$, compute the change of basis matrix from $\mathcal{e}$ to $\mathcal{e}'$. Self-check: pick some $v \in V$, determine $[v]_{\mathcal{e}}$ and $[v]_{\mathcal{e}'}$, and verify that Equation (16.19) holds.

(a) $V = G_2$, $\mathcal{e} = (e_1, e_2)$ is two perpendicular unit vectors, and $\mathcal{e}' = (e_1, e_2')$, where $e_2'$ is $e_1$ rotated by $\theta$.

(b) $V = P_2[\mathbb{F}]$, $\mathcal{e} = (1, t, t^2)$, $\mathcal{e}' = (t^2, (t+1)^2, (t+2)^2)$.

(c) $V = \mathbb{F}^3$, $\mathcal{e} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathcal{e}' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Just as coordinates change with respect to different bases, so do the matrix representations of linear maps.

Proposition 16.6.7. Let $V$ and $W$ be finite dimensional vector spaces, let $\mathcal{e}$ and $\mathcal{e}'$ be bases of $V$, let $f$ and $f'$ be bases of $W$, and let $\varphi : V \to W$ be a linear map. Then

$$[\varphi]_{\text{new}} = T^{-1}[\varphi]_{\text{old}}S$$ \hfill (16.20)

where $S$ is the change of basis matrix from $\mathcal{e}$ to $\mathcal{e}'$ and $T$ is the change of basis matrix from $\mathcal{f}$ to $\mathcal{f}'$.

Proposition 16.6.8. Let $N$ be a nilpotent matrix. Then the linear transformation defined by $x \mapsto Nx$ has a chain of invariant subspaces.

Proposition 16.6.9. Every matrix $A \in M_n(\mathbb{C})$ is similar to a triangular matrix.
Chapter 17

(F) Block Matrices (optional)

17.1 Block matrix basics

Definition 17.1.1 (Block matrix). We sometimes write matrices as block matrices, where each entry actually represents a matrix.

Example 17.1.2. The matrix

\[
A = \begin{pmatrix}
1 & 2 & -3 & 4 \\
2 & 1 & 6 & -1 \\
0 & -3 & 1 & 2 \\
\end{pmatrix}
\]

may be symbolically written as the block matrix

\[
A = \begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22} \\
\end{pmatrix}
\]

where \( A_{11} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \ A_{12} = \begin{pmatrix} -3 & 4 \\ 6 & -1 \end{pmatrix}, \ A_{21} = (0, -3), \) and \( A_{22} = (-1, 2). \)

Definition 17.1.3 (Block-diagonal matrix). A square matrix \( A \) is said to be a block-diagonal matrix if it can be written in the form

\[
A = \begin{pmatrix}
A_1 & 0 & \cdots & 0 \\
0 & A_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & A_k \\
\end{pmatrix}
\]

where the diagonal blocks \( A_i \) are square matrices. In this case, we say that \( A \) is the diagonal sum of the matrices \( A_1, \ldots, A_k. \)

Example 17.1.4. The matrix

\[
A = \begin{pmatrix}
1 & 2 & 0 & 0 & 0 & 0 \\
-3 & 7 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 6 & 0 & 0 \\
0 & 0 & 0 & 2 & -1 & 3 \\
0 & 0 & 0 & 3 & 2 & 5 \\
\end{pmatrix}
\]

is a block-diagonal matrix. We may write

\[
A = \begin{pmatrix}
A_1 & 0 & 0 \\
0 & A_2 & 0 \\
0 & 0 & A_3 \\
\end{pmatrix}
\]

where \( A_1 = \begin{pmatrix} 1 & 2 \\ -3 & 7 \end{pmatrix}, \ A_2 = (6), \) and \( A_3 = \begin{pmatrix} 4 & 6 & 0 \\
2 & -1 & 3 \\
3 & 2 & 5 \end{pmatrix}. \)

Note that the blocks need not be of the same size.

Definition 17.1.5 (Block-triangular matrix). A square matrix \( A \) is said to be a block-triangular matrix if it can be written in the form

\[
A = \begin{pmatrix}
A_{11} & A_{12} & \cdots & A_{1k} \\
A_{21} & A_{22} & \cdots & A_{2k} \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & A_{kk} \\
\end{pmatrix}
\]

for blocks \( A_{ij}, \) where the diagonal blocks \( A_{ii} \) are square matrices.
Exercise 17.1.6. Show that block matrices multiply in the same way as matrices. More precisely, suppose that 
\[ A = (A_{ij}) \] and 
\[ B = (B_{jk}) \] are block matrices where 
\[ A_{ij} \in \mathbb{F}^{r_i \times s_j} \] and 
\[ B_{jk} \in \mathbb{F}^{s_j \times t_k} \]. Let 
\[ C = AB \] (why is this product defined?). Show that 
\[ C = (C_{ik}) \] where 
\[ C_{ik} \in \mathbb{F}^{r_i \times t_k} \] and 
\[ C_{jk} = \sum_j A_{ij} B_{jk} \] . \hspace{1cm} (17.1)

17.2 Arithmetic of block-diagonal and block-triangular matrices

Proposition 17.2.1. Let 
\[ A = \text{diag}(A_1, \ldots, A_n) \] and 
\[ B = \text{diag}(B_1, \ldots, B_n) \] be block-diagonal matrices with blocks of the same size and let 
\[ \lambda \in \mathbb{F} \]. Then
\[
A + B = \text{diag}(A_1 + B_1, \ldots, A_n + B_n) \hspace{1cm} (17.2)
\]
\[
\lambda A = \text{diag}(\lambda A_1, \ldots, \lambda A_n) \hspace{1cm} (17.3)
\]
\[
AB = \text{diag}(A_1 B_1, \ldots, A_n B_n) \hspace{1cm} (17.4)
\]

Proposition 17.2.2. Let 
\[ A = \text{diag}(A_1, \ldots, A_n) \] be a block-diagonal matrix. Then 
\[ A^k = \text{diag}(A_1^k, \ldots, A_n^k) \] for all \( k \).

Proposition 17.2.3. Let \( f \in \mathbb{F}[t] \) be a polynomial and let 
\[ A = \text{diag}(A_1, \ldots, A_n) \] be a block-diagonal matrix. Then
\[
f(A) = \text{diag}(f(A_1), \ldots, f(A_n)) \hspace{1cm} (17.5)
\]

In our discussion of the arithmetic block-triangular matrices, we are interested only in the block-diagonal entries.

Proposition 17.2.4. Let
\[ A = \begin{pmatrix} A_1 & \ast \\ \ast & \ddots \end{pmatrix} \]
and
\[ B = \begin{pmatrix} B_1 & \ast \\ \ast & \ddots \end{pmatrix} \]
be block-upper triangular matrices with blocks of the same size and let \( \lambda \in \mathbb{F} \). Then
\[
A + B = \begin{pmatrix} A_1 + B_1 & \ast \\ \ast & \ddots \end{pmatrix} \hspace{1cm} (17.6)
\]
\[
\lambda A = \begin{pmatrix} \lambda A_1 & \ast \\ \ast & \ddots \end{pmatrix} \hspace{1cm} (17.7)
\]
\[
AB = \begin{pmatrix} A_1 B_1 & \ast \\ \ast & \ddots \end{pmatrix} \hspace{1cm} (17.8)
\]
Proposition 17.2.5. Let $A$ be as in Prop. 17.2.4. Then
\[
A^k = \begin{pmatrix}
A_1^k & \ast \\
A_2^k & \\
0 & \ddots \\
& & & A_n^k
\end{pmatrix}
\] (17.9)
for all $k$.

Proposition 17.2.6. Let $f \in \mathbb{F}[t]$ be a polynomial and let $A$ be as in Prop. 17.2.4. Then
\[
f(A) = \begin{pmatrix}
f(A_1) & \ast \\
f(A_2) & \\
0 & \ddots \\
& & & f(A_n)
\end{pmatrix}.
\] (17.10)
Chapter 18

(F) Minimal Polynomials of Matrices and Linear Transformations (optional)

18.1 The minimal polynomial

All matrices in this section are square.

In Section 8.5, we defined what it means to plug a matrix into a polynomial ([1] Def. 2.3.3).

Definition 18.1.1 (Annihilation). Let \( f \in \mathbb{F}[t] \). We say that \( f \) annihilates the matrix \( A \in M_n(\mathbb{F}) \) if \( f(A) = 0 \).

Exercise 18.1.2. Let \( A \) be a matrix. Prove, without using the Cayley-Hamilton Theorem, that for every matrix there is a nonzero polynomial that annihilates it, i.e., for every matrix \( A \) there is a nonzero polynomial \( f \) such that \( f(A) = 0 \). Show such a polynomial of degree at most \( n^2 \) exists.

Definition 18.1.3 (Minimal polynomial). Let \( A \in M_n(\mathbb{F}) \). The polynomial \( f \in \mathbb{F}[t] \) is a minimal polynomial of \( A \) if \( f \) is a nonzero polynomial of lowest degree that annihilates \( A \).

Example 18.1.4. The polynomial \( t - 1 \) is a minimal polynomial of the \( n \times n \) identity matrix.

Exercise 18.1.5. Let \( A \) be the diagonal matrix \( \text{diag}(3, 3, 7, 7) \), i.e.,

\[
A = \begin{pmatrix} 3 & 0 & & \\ 3 & 7 & & \\ & 0 & 7 & \\ & & & 7 \end{pmatrix}.
\]

Find a minimal polynomial for \( A \). Compare your answer to the characteristic polynomial \( f_A \). Recall ([1] Prop. 2.3.4) that for all \( f \in \mathbb{F}[t] \), we have

\[
f(\text{diag}(\lambda_1, \ldots, \lambda_n)) = \text{diag}(f(\lambda_1), \ldots, f(\lambda_n)).
\]

Proposition 18.1.6. Let \( A \in M_n(\mathbb{F}) \) and let \( m \) be a minimal polynomial of \( A \). Then for all \( g \in \mathbb{F}[t] \), we have \( g(A) = 0 \) if and only if \( m \mid g \).

Corollary 18.1.7. Let \( A \in M_n(\mathbb{F}) \). Then the minimal polynomial of \( A \) is unique up to nonzero scalar factors.
Convention 18.1.8. When discussing “the” minimal polynomial of a matrix, we refer to the unique monic minimal polynomial, denoted \( m_A \).

Corollary 18.1.9 (Cayley-Hamilton re-stated). The minimal polynomial of a matrix divides its characteristic polynomial.

Corollary 18.1.10. Let \( A \in M_n(\mathbb{F}) \). Then \( \deg m_A \leq n \).

Example 18.1.11. \( m_I = t - 1 \).

Exercise 18.1.12. Let \( A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \). Prove \( m_A = (t - 1)^2 \).

Exercise 18.1.13. Find two \( 2 \times 2 \) matrices with the same characteristic polynomial but different minimal polynomials.

Exercise 18.1.14. Let \( A = \text{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_k) \) where the \( \lambda_i \) are distinct. Prove that \( m_A = \prod_{i=1}^{k}(t - \lambda_i) \). \hfill (18.1)

Exercise 18.1.15. Let \( A \) be a block-diagonal matrix, say \( A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \\ & \ddots \\ & & \ddots \\ & & & A_k \end{pmatrix} \). Give a simple expression of \( m_A \) in terms of the \( m_{A_i} \).

Proposition 18.1.16. Let \( A \in M_n(\mathbb{F}) \). Then \( \lambda \) is an eigenvalue of \( A \) if and only if \( m_A(\lambda) = 0 \).

Exercise 18.1.17. Prove: similar matrices have the same minimal polynomial, i.e., if \( A \sim B \) then \( m_A = m_B \).

Proposition 18.1.18. Let \( A \in M_n(\mathbb{C}) \). If \( m_A \) does not have multiple roots then \( A \) is diagonalizable.

In fact, this condition is necessary and sufficient (\[ \text{Theorem 18.2.20} \].)

### 18.2 Minimal polynomials of linear transformations

In this section, \( V \) is an \( n \)-dimensional vector space over \( \mathbb{F} \).

In Section 16.4.2 we defined what it means to plug a linear transformation into a polynomial (\[ \text{Def. 16.4.37} \].)

Exercise 18.2.1. Define what it means for the polynomial \( f \in \mathbb{F}[t] \) to annihilate the linear transformation \( \varphi : V \to V \).

Exercise 18.2.2. Let \( \varphi : V \to V \) be a linear transformation. Prove that there is a nonzero polynomial that annihilates \( \varphi \).

Exercise 18.2.3. Define a minimal polynomial of a linear transformation \( \varphi : V \to V \).
Example 18.2.4. The polynomial \( t - 1 \) is a minimal polynomial of the identity transformation.

Proposition 18.2.5. Let \( \varphi : V \to V \) be a linear transformation and let \( m \) be a minimal polynomial of \( \varphi \). Then for all \( g \in \mathbb{F}[t] \), we have \( g(\varphi) = 0 \) if and only if \( m \mid g \).

Corollary 18.2.6. Let \( \varphi : V \to V \) be a linear transformation. Then the minimal polynomial of \( \varphi \) is unique up to nonzero scalar factors.

Convention 18.2.7. When discussing “the” minimal polynomial of a linear transformation, we shall refer to the unique monic minimal polynomial, denoted \( m_\varphi \).

Proposition 18.2.8. Let \( \varphi : V \to V \) be a linear transformation. Then \( \deg m_\varphi \leq n \).

Exercise 18.2.10. Let \( \varphi : V \to V \) be a linear transformation with an eigenbasis. Prove: all roots of \( m_\varphi \) belong to \( \mathbb{F} \) and \( m_\varphi \) has no multiple roots. Let \( b_1, \ldots, b_n \) be an eigenbasis, and let \( \varphi(b_i) = \lambda_i b_i \) for each \( i \). Find \( m_\varphi \) in terms of the \( \lambda_i \).

Proposition 18.2.11. Let \( \varphi : V \to V \) be a linear transformation. Then \( \lambda \) is an eigenvalue of \( \varphi \) if and only if \( m_\varphi(\lambda) = 0 \).

Exercise 18.2.12 (Consistency of translation). (a) Let \( b \) be a basis of \( V \) and let \( \varphi : V \to V \) be a linear transformation. Let \( A = [\varphi]_b \). Show that \( m_\varphi = m_A \).

(b) Use this to give a second proof of Ex. 18.1.17 (similar matrices have the same minimal polynomial).

Recall the definition of an invariant subspace of the linear transformation \( \varphi \) (Def. 16.4.22).

Definition 18.2.13 (Minimal invariant subspace). Let \( \varphi : V \to V \) and let \( W \leq V \). We say that \( W \) is a minimal \( \varphi \)-invariant subspace if \( W \) is a \( \varphi \)-invariant subspace, \( W \neq \{0\} \), and the only \( \varphi \)-invariant subspaces of \( W \) are \( W \) and \( \{0\} \).

Lemma 18.2.14. Let \( \varphi : V \to V \) be a linear transformation and let \( W \leq V \) be a \( \varphi \)-invariant subspace. Let \( \varphi_W \) denote the restriction of \( \varphi \) to \( W \). Then \( m_{\varphi_W} | m_\varphi \). ♦

Proposition 18.2.15. Let \( V = W_1 \oplus \cdots \oplus W_k \) where the \( W_i \) are \( \varphi \)-invariant subspaces and \( \oplus \) denotes their direct sum (Def. 15.5.2). If \( m_i \) denotes the minimal polynomial of \( \varphi_{W_i} \) then

\[
m_\varphi = \text{lcm} m_i . \quad (18.2)
\]

(a) Prove this without using Ex. 18.1.15, i.e., without translating the problem to matrices.

(b) Prove this using Ex. 18.1.15.
Theorem 18.2.16. Let $V$ be a finite-dimensional vector space and let $\varphi : V \to V$ be a linear transformation. If $m_\varphi$ has an irreducible factor $f$ of degree $d$, then there is a $d$-dimensional minimal invariant subspace $W$ such that
\[ m_\varphi W = f W = f. \] (18.3)

Corollary 18.2.17. Let $V$ be a finite-dimensional vector space over $\mathbb{R}$ and let $\varphi : V \to V$ be a linear transformation. Then $V$ has a $\varphi$-invariant subspace of dimension 1 or 2.

Proposition 18.2.18. Let $\varphi : V \to V$ be a linear transformation. Let $f \mid m_\varphi$ and let $W = \ker f(\varphi)$. Then

(a) $m_\varphi W \mid f$;

(b) if $\gcd\left(f, \frac{m_\varphi}{f}\right) = 1$, then $m_\varphi W = f$.

(a) $f(\varphi W) = f(\varphi)W = 0$.

(b) $m_\varphi = \operatorname{lcm}(m_\varphi W, m_\varphi W')$ where $W' = \ker \frac{m_\varphi}{f}(\varphi)$, i.e., $V = W \oplus W'$.

Proposition 18.2.19. Let $\varphi : V \to V$ be a linear transformation and let $f, g \in \mathbb{F}[t]$ with $m_\varphi = f \cdot g$ and $\gcd(f, g) = 1$. Then
\[ \mathbb{F}^n = (\ker f(\varphi)) \oplus (\ker g(\varphi)) \] (18.4)

Theorem 18.2.20. Let $A \in M_n(\mathbb{C})$. Then $A$ is diagonalizable if and only if $m_A$ does not have multiple roots over $\mathbb{C}$. ♦

Theorem 18.2.21. Let $\mathbb{F} = \mathbb{C}$ and let $\varphi : V \to V$ be a linear transformation. Then $\varphi$ has an eigenbasis if and only if $m_\varphi$ does not have multiple roots. ♦

The next theorem represents a step toward canonical forms of matrices.

Theorem 18.2.22. Let $A \in M_n(\mathbb{F})$. Then there is a matrix $B \in M_n(\mathbb{F})$ such that $A \sim B$ and $B$ is the diagonal sum of matrices whose minimal polynomials are powers of irreducible polynomials. ♦
Chapter 19

(R) Euclidean Spaces

19.1 Inner products

Let $V$ be a vector space over $\mathbb{R}$.

In Section 1.4, we introduced the standard dot product of $\mathbb{R}^n$ (Def. 1.4.1). We now generalize this to the notion of an inner product over a real vector space $V$.

**Definition 19.1.1 (Euclidean space).** A Euclidean space $V$ is a vector space over $\mathbb{R}$ endowed with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ which is positive definite, symmetric, and bilinear. That is, for all $u, v, w \in V$ and $\alpha \in \mathbb{R}$, we have

(a) $\langle v, v \rangle \geq 0$, with equality holding if and only if $v = 0$ (positive definite)

(b) $\langle v, w \rangle = \langle w, v \rangle$ (symmetric)

(c) $\langle v, \alpha u + w \rangle = \langle v, w \rangle + \alpha \langle v, u \rangle$ (bilinear)

Observe that the standard dot product of $\mathbb{R}^n$ that was introduced in Section 1.4 has all of these properties and, in particular, the vector space $\mathbb{R}^n$ endowed with this inner product is a Euclidean space.

**Exercise 19.1.2.** Let $V$ be a Euclidean space with inner product $\langle \cdot, \cdot \rangle$. Show that for all $v \in V$, we have $\langle v, 0 \rangle = \langle 0, v \rangle = 0$.

**Examples 19.1.3.** The following vector spaces together with the specified inner products are Euclidean spaces (verify this).

(a) $V = \mathbb{R}[t], \langle f, g \rangle = \int_{-\infty}^{\infty} f(t)g(t)dt$ where $\rho(t)$ is a nonnegative continuous function which is not identically 0 and has the property that $\int_{-\infty}^{\infty} \rho(t)t^{2n}dt < \infty$ for all nonnegative integers $n$ (such a function $\rho$ is called a weight function)

(b) $V = C[0, 1]$ (the space of continuous functions $f : [0, 1] \to \mathbb{R}$) with $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$

(c) $V = \mathbb{R}^{k \times n}, \langle A, B \rangle = \text{Tr} \left( AB^T \right)$

(d) $V = \mathbb{R}^n$, and $\langle x, y \rangle = x^TAy$ where $A \in M_n(\mathbb{R})$ is a symmetric positive definite (Def. 10.2.3) $n \times n$ real matrix

Notice that the same vector space can be endowed with different inner products (for example, different weight functions for the inner product of functions).
product on \( \mathbb{R}[t] \), so that there are many Euclidean spaces with the same underlying vector space.

Because they have inner products, Euclidean spaces carry with them the notion of distance (“norm”) and the notion of two vectors being perpendicular (“orthogonality”). Just as inner products generalize the standard dot product in \( \mathbb{R}^n \), these concepts generalize the definitions of norm and orthogonality presented for \( \mathbb{R}^n \) (with respect to the standard dot product) in Section 1.4.

**Definition 19.1.4 (Norm).** Let \( V \) be a Euclidean space, and let \( v \in V \). Then the norm of \( v \), denoted \( \|v\| \), is

\[
\|v\| := \sqrt{\langle v, v \rangle} .
\]  (19.2)

The notion of a norm allows us to easily define the distance between two vectors.

**Definition 19.1.5.** Let \( V \) be a Euclidean space, and let \( v, w \in V \). Then the distance between the vectors \( v \) and \( w \), denoted \( d(v, w) \), is

\[
d(v, w) := \|v - w\| .
\]  (19.3)

The following two theorems show that distance in Euclidean spaces behaves the way we are used to it behaving in \( \mathbb{R}^n \).

**Theorem 19.1.6 (Cauchy-Schwarz inequality).** Let \( V \) be a Euclidean space, and let \( v, w \in V \). Then

\[
\|v + w\| \leq \|v\| + \|w\| .
\]  (19.5)

\( \lozenge \)

**Exercise 19.1.8.** Show that the triangle inequality is equivalent to the Cauchy-Schwarz inequality.

**Definition 19.1.9 (Angle between vectors).** Let \( V \) be a Euclidean space, and let \( v, w \in V \). The angle \( \theta \) between \( v \) and \( w \) is defined by

\[
\theta := \arccos \frac{\langle v, w \rangle}{\|v\|\|w\|} .
\]  (19.6)

Observe that from the Cauchy-Schwarz inequality, we have (for \( v, w \neq 0 \)) that

\[
-1 \leq \frac{\langle v, w \rangle}{\|v\|\|w\|} \leq 1
\]  (19.7)

so this definition is valid, as this term is always in the domain of \( \arccos \).

**Definition 19.1.10 (Orthogonality).** Let \( V \) be a Euclidean space, and let \( v, w \in V \). Then \( v \) and \( w \) are orthogonal (denoted \( v \perp w \)) if \( \langle v, w \rangle = 0 \). A set of vectors \( S = \{v_1, \ldots, v_n\} \) is said to be orthogonal if \( v_i \perp v_j \) whenever \( i \neq j \).

Observe that two vectors are orthogonal when the angle between them is \( \frac{\pi}{2} \). This agrees with the notion of orthogonality being a generalization of perpendicularity.
Exercise 19.1.11. Let $V$ be a Euclidean space. What vectors are orthogonal to every vector?

Exercise 19.1.12. Let $V = C[0, 2\pi]$ be the space of continuous functions $f : [0, 2\pi] \to \mathbb{R}$, endowed with the inner product
\[
\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt . \tag{19.8}
\]

Show that the set \{1, $\cos t$, $\sin t$, $\cos(2t)$, $\sin(2t)$, $\cos(3t)$, $\ldots$\} is an orthogonal set in this Euclidean space.

Definition 19.1.13 (Orthogonal system). An orthogonal system in a Euclidean space $V$ is a list of (pairwise) orthogonal nonzero vectors in $V$.

Proposition 19.1.14. Every orthogonal system in a Euclidean space is linearly independent.

Definition 19.1.15 (Gram matrix). Let $V$ be a Euclidean space, and let $v_1, \ldots, v_k \in V$. The Gram matrix of $v_1, \ldots, v_k$ is the $k \times k$ matrix whose $(i, j)$ entry is $\langle v_i, v_j \rangle$, that is,
\[
G = G(v_1, \ldots, v_k) := (\langle v_i, v_j \rangle)_{i,j=1}^k . \tag{19.9}
\]

Exercise 19.1.16. Let $V$ be a Euclidean space. Show that the vectors $v_1, \ldots, v_k \in V$ are linearly independent if and only if $\det G(v_1, \ldots, v_k) \neq 0$.

Exercise 19.1.17. Let $V$ be a Euclidean space and let $v_1, \ldots, v_k \in V$. Show
\[
\text{rk}(v_1, \ldots, v_k) = \text{rk}(G(v_1, \ldots, v_k)) . \tag{19.10}
\]

Definition 19.1.18 (Orthonormal system). An orthonormal system in a Euclidean space $V$ is a list of (pairwise) orthogonal vectors in $V$, all of which have unit norm. So $(v_1, v_2, \ldots)$ is an orthonormal system if $\langle v_i, v_j \rangle = \delta_{ij}$ for all $i, j$.

In the case of finite-dimensional Euclidean spaces, we are particularly interested in orthonormal bases.

Definition 19.1.19 (Orthonormal basis). Let $V$ be a Euclidean space. A list $(v_1, \ldots, v_n)$ is an orthonormal basis (ONB) of $V$ if it is a basis and $\{v_1, \ldots, v_n\}$ is an orthonormal set.

In the next section, not only will we show that every finite-dimensional Euclidean space has an orthonormal basis, but we will, moreover, demonstrate an algorithm for transforming any basis into an orthonormal basis.

Proposition 19.1.20. Let $V$ be a Euclidean space with orthonormal basis $b$. Then for all $v, w \in V$,
\[
\langle v, w \rangle = [v]_b^T[w]_b . \tag{19.11}
\]

Proposition 19.1.21. Let $V$ be a Euclidean space. Every linear form $f : V \to \mathbb{R}$ (Def. 15.1.6) can be written as
\[
f(x) = \langle a, x \rangle \tag{19.12}
\]
for a unique $a \in V$. 

19.2 Gram-Schmidt orthogonalization

The main result of this section is the following theorem.

**Theorem 19.2.1.** Every finite-dimensional Euclidean space has an orthonormal basis. In fact, every orthonormal system extends to an orthonormal basis. ♦

Before we prove this theorem, we will first develop *Gram-Schmidt orthogonalization*, an online procedure that takes a list of vectors as input and produces a list of orthogonal vectors satisfying certain conditions. We formalize this below.

**Theorem 19.2.2 (Gram-Schmidt orthogonalization).** Let $V$ be a Euclidean space and let $v_1, \ldots, v_k \in V$. For $i = 1, \ldots, k$, let $U_i = \text{span}(v_1, \ldots, v_i)$. Then there exist vectors $e_1, \ldots, e_k$ such that

(a) For all $j \geq 1$, we have $v_j - e_j \in U_{j-1}$

(b) The $e_j$ are pairwise orthogonal

Moreover, the $e_j$ are uniquely determined. ♦

Note that $U_0 = \text{span}(\emptyset) = \{0\}$.

Let $\text{GS}(k)$ denote the statement of Theorem 19.2.2 for a particular value of $k$. We prove the Theorem 19.2.2 by induction on $k$. The inductive step is based on the following lemma.

**Lemma 19.2.3.** Assume $\text{GS}(k)$ holds. Then, for all $j \leq k$, we have $\text{span}(e_1, \ldots, e_j) = U_j$. ♦

**Exercise 19.2.4.** Prove $\text{GS}(1)$.

**Exercise 19.2.5.** Let $k \geq 2$ and assume $\text{GS}(k-1)$ is true. Look for $e_k$ in the form

$$e_k = v_k - \sum_{i=1}^{k-1} \alpha_i e_i$$

Prove that the only possible vector $e_k$ satisfying the conditions of $\text{GS}(k)$ is the $e_k$ for which

$$\alpha_i = \frac{\langle v_i, e_i \rangle}{\|e_i\|^2}$$

except in the case where $e_i = 0$ (in that case, $\alpha_i$ can be chosen arbitrarily).

**Exercise 19.2.6.** Prove that $e_k$ as constructed in the previous exercise satisfies $\text{GS}(k)$.

This completes the proof of Theorem 19.2.2.

**Proposition 19.2.7.** $e_i = 0$ if and only if $v_i \in \text{span}(v_1, \ldots, v_{i-1})$.

**Proposition 19.2.8.** The Gram-Schmidt procedure preserves linear independence.

**Proposition 19.2.9.** If $(v_1, \ldots, v_k)$ is a basis of $V$, then so is $(e_1, \ldots, e_k)$.

**Exercise 19.2.10.** Let $\mathcal{E} = (e_1, \ldots, e_n)$ be an orthonormal basis of $V$. From $\mathcal{E}$, construct an orthonormal basis $\mathcal{E}' = (e'_1, \ldots, e'_n)$ of $V$.

**Exercise 19.2.11.** Conclude that every finite-dimensional Euclidean space has an orthonormal basis.
Numerical exercise 19.2.12. Apply the Gram-Schmidt procedure to find an orthonormal basis for each of the following Euclidean spaces $V$ from the basis $b$. Self-check: once you have applied Gram-Schmidt, verify that you have obtained an orthonormal set of vectors.

(a) $V = \mathbb{R}^3$ with the standard dot product, $b = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}$

(b) $V = P_2[\mathbb{R}]$, with $\langle f, g \rangle = \int_{-\infty}^{\infty} e^{-t^2/2} f(t)g(t)dt$, and $b = (1, t, t^2)$ (Hermite polynomials)

(c) $V = P_2[\mathbb{R}]$, with $\langle f, g \rangle = \int_{-\infty}^{\infty} \frac{1}{\sqrt{1-t^2}} f(t)g(t)dt$, and $b = (1, t, t^2)$ (Chebyshev polynomials of the first kind)

(d) $V = P_2[\mathbb{R}]$, with $\langle f, g \rangle = \int_{-\infty}^{\infty} \sqrt{1-t^2} f(t)g(t)dt$, and $b = (1, t, t^2)$ (Chebyshev polynomials of the second kind)

19.3 Isometries and Orthogonality

Definition 19.3.1 (Isometry). Let $V$ and $W$ be Euclidean spaces. Then an isometry $f : V \to W$ is a bijection that preserves inner product, i.e., for all $v_1, v_2 \in V$, we have

$$\langle v_1, v_2 \rangle_V = \langle f(v_1), f(v_2) \rangle_W$$  \hspace{1cm} (19.15)

The Euclidean spaces $V$ and $W$ are isometric if there is an isometry between them.

Theorem 19.3.2. If $V$ and $W$ are finite dimensional Euclidean spaces, then they are isometric if and only if $\dim V = \dim W$. ♦

Proposition 19.3.3. Let $V$ and $W$ be Euclidean spaces. Then $\varphi : V \to W$ is an isometry if and only if it maps an orthonormal basis of $V$ to an orthonormal basis of $W$.

Proposition 19.3.4. Let $\varphi : V \to W$ be an isomorphism that preserves orthogonality (so $v \perp w$ if and only if $\varphi(v) \perp \varphi(w)$). Show that there is an isometry $\psi$ and a nonzero scalar $\lambda$ such that $\varphi = \lambda \psi$.

The geometric notion of congruence is captured by the concept of orthogonal transformations.

Definition 19.3.5 (Orthogonal transformation). Let $V$ be a Euclidean space. A linear transformation $\phi : V \to V$ is called an orthogonal transformation if it is an isometry. The set of orthogonal transformations of $V$ is denoted by $O(V)$, and is called the orthogonal group of $V$.

Proposition 19.3.6. The set $O(V)$ is a group (Def. 14.2.1) under composition.

Exercise 19.3.7. The linear transformation $\varphi : V \to V$ is orthogonal if and only if $\varphi$ preserves the norm, i.e., for all $v \in \mathbb{C}^n$, we have $\|\varphi v\| = \|v\|$.

Theorem 19.3.8. Let $\varphi \in O(V)$. Then all eigenvalues of $\varphi$ are $\pm 1$. ♦
Proposition 19.3.9. Let $V$ be a Euclidean space and let $e = (e_1, \ldots, e_n)$ be an orthonormal basis of $V$. Then $\varphi : V \to V$ is an orthogonal transformation if and only if $(\varphi(e_1), \ldots, \varphi(e_n))$ is an orthonormal basis.

Proposition 19.3.10 (Consistency of translation). Let $V$ be a Euclidean space with orthonormal basis $b$, and let $\varphi : V \to V$ be a linear transformation. Then $\varphi$ is orthogonal if and only if $[\varphi]_b$ is an orthogonal matrix (Def. 9.1.1).

Definition 19.3.11. Let $V$ be a Euclidean space and let $S, T \subseteq V$. For $v \in V$, we say that $v$ is orthogonal to $S$ (notation: $v \perp S$) if for all $s \in S$, we have $v \perp s$. Moreover, we say that $S$ is orthogonal to $T$ (notation: $S \perp T$) if $s \perp t$ for all $s \in S$ and $t \in T$.

Definition 19.3.12. Let $V$ be a Euclidean space and let $S \subseteq V$. Then $S^\perp$ ("$S$ perp") is the set of vectors orthogonal to $S$, i.e.,

$$S^\perp := \{ v \in V \mid v \perp S \}. \quad (19.16)$$

Proposition 19.3.13. For all subsets $S \subseteq V$, we have $S^\perp \subseteq V$.

Proposition 19.3.14. Let $S \subseteq V$. Then $S \subseteq (S^\perp)^\perp$.

Exercise 19.3.15. Verify

(a) $\{0\}^\perp = V$

(b) $\emptyset^\perp = V$

(c) $V^\perp = \{0\}$

The next theorem says that the direct sum (Def. 15.5.2) of a subspace and its perp is the entire space.

Theorem 19.3.16. If $\dim V < \infty$ and $W \leq V$, then $V = W \oplus W^\perp$.

The proof of this theorem requires the following lemma.

Lemma 19.3.17. Let $V$ be a vector space with $\dim V = n$, and let $W \leq V$. Then

$$\dim W^\perp = n - \dim W. \quad (19.17)$$

Proposition 19.3.18. Let $V$ be a finite-dimensional Euclidean space and let $S \subseteq V$. Then

$$(S^\perp)^\perp = \text{span}(S). \quad (19.18)$$

Proposition 19.3.19. Let $U_1, \ldots, U_k$ be pairwise orthogonal subspaces (i.e., $U_i \perp U_j$ whenever $i \neq j$). Then

$$\sum_{i=1}^k U_i = \bigoplus_{i=1}^k U_i. \quad (19.19)$$

We now study the linear map analogue of the transpose of a matrix, known as the adjoint of the linear map.

Theorem 19.3.20. Let $V$ and $W$ be Euclidean spaces, and let $\varphi : V \to W$ be a linear map. Then there exists a unique linear map $\psi : W \to V$ such that for all $v \in V$ and $w \in W$, we have

$$\langle \varphi v, w \rangle = \langle v, \psi w \rangle. \quad (19.19)$$
Note that the inner product above refers to inner products in two different spaces. To be more specific, we should have written
\[ \langle \varphi v, w \rangle_W = \langle v, \psi w \rangle_V. \]
(19.20)

**Definition** 19.3.21. The linear map \( \psi \) whose existence is guaranteed by Theorem 19.3.20 is called the adjoint of \( \varphi \) and is denoted \( \varphi^* \). So for all \( v \in V \) and \( w \in W \), we have
\[ \langle \varphi v, w \rangle = \langle v, \varphi^* w \rangle. \]
(19.21)

The next exercise shows the relationship between the coordinatization of \( \varphi \) and of \( \varphi^* \). The reason we denote the adjoint of the linear map \( \varphi \) by \( \varphi^* \) rather than by \( \varphi^T \) will become clear in Section 20.4.

**Proposition 19.3.22.** Let \( V, W \), and \( \varphi \) be as in the statement of Theorem 19.3.20. Let \( b_1 \) be an orthonormal basis of \( V \) and let \( b_2 \) be an orthonormal basis of \( W \). Then
\[ [\varphi^*]_{b_2, b_1} = [\varphi]^T_{b_1, b_2}. \]
(19.22)

### 19.4 First proof of the Spectral Theorem

We first stated the Spectral Theorem for real symmetric matrices in Chapter 10. We now restate the theorem in the context of Euclidean spaces and break its proof into a series of exercises.

Let \( V \) be a Euclidean space.

**Definition** 19.4.1 (Symmetric linear transformation). Let \( \varphi : V \to V \) be a linear transformation. Then \( \varphi \) is symmetric if, for all \( v, w \in V \), we have
\[ \langle v, \varphi(w) \rangle = \langle \varphi(v), w \rangle. \]
(19.23)

**Proposition 19.4.2.** Let \( \varphi : V \to V \) be a symmetric transformation, and let \( v_1, \ldots, v_k \) be eigenvectors of \( \varphi \) with distinct eigenvalues. Then \( v_i \perp v_j \) whenever \( i \neq j \).

**Proposition 19.4.3.** Let \( b \) be an orthonormal basis of the finite-dimensional Euclidean space \( V \), and let \( \varphi : V \to V \) be a linear transformation. Then \( \varphi \) is symmetric if and only if the matrix \([\varphi]_b\) is symmetric.

In particular, this proposition establishes the equivalence of Theorem 10.1.1 and Theorem 19.4.4.

The main theorem of this section is the Spectral Theorem.

**Theorem 19.4.4 (Spectral Theorem).** Let \( V \) be a finite-dimensional Euclidean space and let \( \varphi : V \to V \) be a symmetric linear transformation. Then \( \varphi \) has an orthonormal eigenbasis.

**Proposition 19.4.5.** Let \( \varphi \) be a symmetric linear transformation of the Euclidean space \( V \), and let \( W \leq V \). If \( W \) is \( \varphi \)-invariant (Def. 16.4.22) then \( W^\perp \) is \( \varphi \)-invariant.

**Lemma 19.4.6.** Let \( \varphi : V \to V \) be a symmetric linear transformation and let \( W \) be a \( \varphi \)-invariant subspace of \( V \). Then the restriction of \( \varphi \) to \( W \) is also symmetric.
The heart of the proof of the Spectral Theorem is the following lemma.

**Main Lemma 19.4.7.** Let \( \varphi \) be a symmetric linear transformation of a Euclidean space of degree \( \geq 1 \). Then \( \varphi \) has an eigenvector.

**Exercise 19.4.8.** Assuming Lemma 19.4.7, prove the Spectral Theorem by induction on \( \dim V \).

Before proving the Main Lemma, we digress briefly into analysis.

**Definition 19.4.9 (Convergence).** Let \( V \) be a Euclidean space and let \( v = (v_1, v_2, v_3, \ldots) \) be a sequence of vectors in \( V \). We say that the sequence \( v \) converges to the vector \( v \) if \( \lim_{i \to \infty} \|v_i - v\| = 0 \), that is, if for all \( \varepsilon > 0 \) there exists \( N \in \mathbb{Z} \) such that \( \|v_i - v\| < \varepsilon \) for all \( i > N \).

**Definition 19.4.10 (Open set).** Let \( S \subseteq \mathbb{R}^n \). We say that \( S \) is open if, for every \( v \in S \), there exists \( \varepsilon > 0 \) such that the set \( \{w \in \mathbb{R}^n \mid \|v - w\| < \varepsilon\} \) is a subset of \( S \).

**Definition 19.4.11 (Closed set).** A set \( S \subseteq V \) is closed if, whenever there is a sequence of vectors \( v_1, v_2, v_3, \ldots \in S \) that converges to a vector \( v \in V \), we have \( v \in S \).

**Definition 19.4.12 (Bounded set).** A set \( S \subseteq V \) is bounded if for all \( w \in V \) and \( r \in \mathbb{R} \) such that for all \( w \in V \), we have \( d(v, w) < r \) (Def. 19.1.5), \( d(v, w) = \|v - w\| \).

**Definition 19.4.13 (Compactness).** Let \( V \) be a finite-dimensional Euclidean space, and let \( S \subseteq V \). Then \( S \) is compact if it is closed and bounded.

**Definition 19.4.14.** Let \( S \) be a set and let \( f : S \to \mathbb{R} \) be a function. We say that \( f \) attains its maximum if there is some \( s_0 \in S \) such that for all \( s \in S \), we have \( f(s_0) \geq f(s) \). The notion of a function attaining its minimum is defined analogously.

**Theorem 19.4.15.** Let \( V \) be a Euclidean space and let \( S \subseteq V \) be a compact and nonempty subset. If \( f : S \to \mathbb{R} \) is continuous, then \( f \) attains its maximum and its minimum.

**Definition 19.4.16 (Rayleigh quotient for linear transformations).** Let \( \varphi : V \to V \) be a linear transformation. Then the Rayleigh quotient of \( \varphi \) is the function \( R_{\varphi} : V \setminus \{0\} \to \mathbb{R} \) defined by

\[
R_{\varphi}(v) := \frac{\langle v, \varphi(v) \rangle}{\|v\|^2}. \tag{19.24}
\]

While this function is defined for all linear transformations, its significance is in its applications to symmetric linear transformations.

**Exercise 19.4.17.** Let \( \varphi : V \to V \) be a linear transformation, and let \( v \in V \) and \( \lambda \in \mathbb{R}^{\times} \). Verify that \( R_{\varphi}(\lambda v) = R_{\varphi}(v) \).

**Proposition 19.4.18.** Let \( \varphi : V \to V \) be a linear transformation. Then the Rayleigh quotient \( R_{\varphi} \) attains its maximum and its minimum.
Exercise 19.4.19. Consider the real function
\[ f(t) = \frac{\alpha + \beta t + \gamma t^2}{\delta + \varepsilon t^2} \]
where \( \alpha, \beta, \gamma, \delta, \varepsilon \in \mathbb{R} \) and \( \delta^2 + \varepsilon^2 \neq 0 \). Suppose \( f(0) \geq f(t) \) for all \( t \in \mathbb{R} \). Show that \( \beta = 0 \).

Definition 19.4.20 (arg max). Let \( S \) be a set and let \( f : S \to \mathbb{R} \) be a function which attains its maximum. Then \( \text{arg max } f := \{ s_0 \in S \mid f(s_0) \geq f(s) \text{ for all } s \in S \} \).

Convention 19.4.21. Let \( S \) be a set and let \( f : S \to \mathbb{R} \) be a function which attains its maximum. We often write \( \text{arg max } f \) to refer to any (arbitrarily chosen) element of the set \( \text{arg max } f \), rather than the set itself.

Proposition 19.4.22. Let \( \varphi \) be a symmetric linear transformation of the Euclidean space \( V \), and let \( v_0 = \text{arg max } R_\varphi(v) \). Then \( v_0 \) is an eigenvector of \( \varphi \).

Prop. 19.4.22 completes the proof of the Main Lemma and thereby the proof of the Spectral Theorem.

Proposition 19.4.23. If two symmetric matrices are similar then they are orthogonally similar.
Chapter 20

(ℂ) Hermitian Spaces

In Chapter 19 we discussed Euclidean spaces, whose underlying vector spaces were real. We now generalize this to the notion of Hermitian spaces, whose underlying vector spaces are complex.

20.1 Hermitian spaces

Definition 20.1.1 (Sesquilinear forms). Let $V$ be a vector space over $ℂ$. The function $f : V \times V \to ℂ$ is a sesquilinear form if the following four conditions are met.

(a) $f(v, w_1 + w_2) = f(v, w_1) + f(v, w_2)$ for all $v, w_1, w_2 \in V$,

(b) $f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w)$ for all $v_1, v_2, w \in V$,

(c) $f(v, \lambda w) = \lambda f(v, w)$ for all $v, w \in V$ and $\lambda \in ℂ$,

(d) $f(\lambda v, w) = \overline{\lambda} f(v, w)$ for all $v, w \in V$ and $\lambda \in ℂ$, where $\overline{\lambda}$ is the complex conjugate of $\lambda$.

Examples 20.1.2. The function $f(v, w) = v^* w$ is a sesquilinear form over $ℂ^n$. More generally, for any $A \in M_n(ℂ)$, $f(v, w) = v^* A w$ (20.1) is sesquilinear.

Exercise 20.1.3. Let $V$ be a vector space over $ℂ$ and let $f : V \times V \to ℂ$ be a sesquilinear form. Show that for all $v \in V$, we have $f(v, 0) = f(0, v) = 0$ (20.2).

Definition 20.1.4 (Hermitian form). Let $V$ be a complex vector space. The function $f : V \times V \to ℂ$ is Hermitian if for all $v, w \in V$, $f(v, w) = \overline{f(w, v)}$ (20.3).

A sesquilinear form that is Hermitian is called a Hermitian form.

Exercise 20.1.5. For what matrices $A \in M_n(ℂ)$ is the sesquilinear form $f(v, w) = v^* A w$ Hermitian?

Exercise 20.1.6. Show that for Hermitian forms, (b) follows from (a) and (d) follows from (c) in Def. 20.1.1.

Fact 20.1.7. Let $V$ be a vector space over $ℂ$ and let $f : V \times V \to ℂ$ be a Hermitian form. Then $f(v, v) \in ℜ$ for all $v \in V$. 

Babai: Discover Linear Algebra. c⃝ 2016 László Babai. 134 This chapter last updated August 9, 2016
20.1. HERMITIAN SPACES

Definition 20.1.8. Let $V$ be a vector space over $\mathbb{C}$, and let $f : V \times V \to \mathbb{C}$ be a Hermitian form.

(a) If $f(v,v) > 0$ for all $v \neq 0$, then $f$ is positive definite.

(b) If $f(v,v) \geq 0$ for all $v \in V$, then $f$ is positive semidefinite.

(c) If $f(v,v) < 0$ for all $v \neq 0$, then $f$ is negative definite.

(d) If $f(v,v) \leq 0$ for all $v \in V$, then $f$ is negative semidefinite.

(e) If there exists $v,w \in V$ such that $f(v,v) > 0$ and $f(w,w) < 0$, then $f$ is indefinite.

Exercise 20.1.9. For what $A \in M_n(\mathbb{C})$ is $v^*Av$ positive definite, positive semidefinite, etc.?

Exercise 20.1.10. Let $V$ be the space of continuous functions $f : [0,1] \to \mathbb{C}$, and let $\rho : [0,1] \to \mathbb{C}$ be a continuous “weight function.” Define

$$F(f,g) := \int_0^1 f(t)\overline{g(t)}\rho(t)dt.$$  \hfill (20.4)

(a) Show that $F$ is a sesquilinear form.

(b) Under what conditions on $\rho$ is $F$ Hermitian?

(c) Under what conditions on $\rho$ is $F$ Hermitian and positive definite?

Exercise 20.1.11. Let $\underline{r} = (\rho_0, \rho_1, \ldots)$ be an infinite sequence of complex numbers. Consider the space $V$ of infinite sequences $(\alpha_0, \alpha_1, \ldots)$ of complex numbers such that

$$\sum_{i=0}^{\infty} |\alpha_i|^2|\rho_i| < \infty.$$ \hfill (20.5)

For $a = (\alpha_0, \alpha_1, \ldots)$ and $b = (\beta_0, \beta_1, \ldots)$, define

$$F(a,b) := \sum_{i=0}^{\infty} \overline{\alpha_i}\beta_i\rho_i.$$ \hfill (20.6)

(a) Show that $F$ is a sesquilinear form.

(b) Under what conditions on $\underline{r}$ is $F$ Hermitian?

(c) Under what conditions on $\underline{r}$ is $F$ Hermitian and positive definite?

Definition 20.1.12 (Hermitian space). A Hermitian space is a vector space $V$ over $\mathbb{C}$ endowed with a positive definite, Hermitian, sesquilinear inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$.

Example 20.1.13. The standard example of a (complex) Hermitian space is $\mathbb{C}^n$ endowed with the standard Hermitian dot product ($\mathbb{C}^n$ Def. [12.2.6]), that is, $\langle v, w \rangle := v^*w$.

Note that the standard Hermitian dot product is sesquilinear and positive definite.

Example 20.1.14. Let $V$ be the space of continuous functions $f : [0,1] \to \mathbb{C}$, and define the inner product

$$\langle f, g \rangle = \int_0^1 \overline{f(t)}g(t)dt$$
Then $V$ endowed with this inner product is a Hermitian space.

**Example 20.1.15.** The space $\ell^2(\mathbb{C})$ of sequences $(\alpha_0, \alpha_1, \ldots)$ of complex numbers such that
\[
\sum_{i=0}^{\infty} |\alpha_i|^2 < \infty
\] (20.7)
endowed with the inner product
\[
\langle a, b \rangle = \sum_{i=0}^{\infty} \overline{a_i}b_i
\] (20.8)
where $a = (\alpha_0, \alpha_1, \ldots)$ and $\overline{b} = (\overline{\beta_0}, \overline{\beta_1}, \ldots)$ is a Hermitian space. This is one of the standard representations of the complex “separable Hilbert space.”

Euclidean spaces generalize the geometric concepts of distance and perpendicularity via the notions of norm and orthogonality, respectively; these are easily extended to complex Hermitian spaces.

**Definition 20.1.16 (Norm).** Let $V$ be a Hermitian space, and let $v \in V$. Then the norm of $v$, denoted $\|v\|$, is defined to be
\[
\|v\| := \sqrt{\langle v, v \rangle}.
\] (20.9)

Just as in Euclidean spaces, the notion of a norm allows us to define the distance between two vectors in a Hermitian space.

**Definition 20.1.17.** Let $V$ be a Hermitian space, and let $v, w \in V$. Then the distance between the vectors $v$ and $w$, denoted $d(v, w)$, is
\[
d(v, w) := \|v - w\|.
\] (20.10)

Distance in Hermitian spaces obeys the same properties that we are used to in Euclidean spaces.

**Theorem 20.1.18 (Cauchy-Schwarz inequality).** Let $V$ be a Hermitian space, and let $v, w \in V$. Then
\[
|\langle v, w \rangle| \leq \|v\| \cdot \|w\|.
\] (20.11)

**Theorem 20.1.19 (Triangle inequality).** Let $V$ be a Hermitian space, and let $v, w \in V$. Then
\[
\|v + w\| \leq \|v\| + \|w\|.
\] (20.12)

Again, like in Euclidean spaces, norms carry with them the notion of angle; however, because $\langle v, w \rangle$ is not necessarily real, our definition of angle is not identical to the definition of angle presented in Section 19.1.

**Definition 20.1.20 (Orthogonality).** Let $V$ be a Hermitian space. Then we say that $v, w \in V$ are orthogonal (notation: $v \perp w$) if $\langle v, w \rangle = 0$.

**Exercise 20.1.21.** Let $V$ be a Hermitian space. What vectors are orthogonal to every vector?

**Definition 20.1.22 (Orthogonal system).** An orthogonal system in a Hermitian space $V$ is a list of (pairwise) orthogonal nonzero vectors in $V$. 
Proposition 20.1.23. Every orthogonal system in a Hermitian space is linearly independent.

Definition 20.1.24 (Gram matrix). Let $V$ be a Hermitian space, and let $v_1, \ldots, v_k \in V$. The Gram matrix of $v_1, \ldots, v_k$ is the $k \times k$ matrix whose $(i, j)$ entry is $\langle v_i, v_j \rangle$, that is,

$$G = G(v_1, \ldots, v_k) := (\langle v_i, v_j \rangle)_{i,j=1}^k. \quad (20.13)$$

Exercise 20.1.25. Let $V$ be a Hermitian space. Show that the vectors $v_1, \ldots, v_k \in V$ are linearly independent if and only if $\det G(v_1, \ldots, v_k) \neq 0$.

Exercise 20.1.26. Let $V$ be a Hermitian space and let $v_1, \ldots, v_k \in V$. Show

$$\text{rk}(v_1, \ldots, v_k) = \text{rk}(G(v_1, \ldots, v_k)). \quad (20.14)$$

Definition 20.1.27 (Orthonormal system). An orthonormal system in a Hermitian space $V$ is a list of (pairwise) orthogonal vectors in $V$, all of which have unit norm. So $(v_1, v_2, \ldots)$ is an orthonormal system if $\langle v_i, v_j \rangle = \delta_{ij}$ for all $i, j$.

In the case of finite-dimensional Hermitian spaces, we are particularly interested in orthonormal bases, just as we were interested in orthonormal bases for finite-dimensional Euclidean spaces.

Definition 20.1.28 (Orthonormal basis). Let $V$ be a Hermitian space. An orthonormal basis of $V$ is an orthonormal system that is a basis of $V$.

Exercise 20.1.29. Generalize the Gram-Schmidt orthogonalization procedure (Section 19.2) to complex Hermitian spaces. Theorem 19.2.2 will hold verbatim, replacing the word “Euclidean” by “Hermitian.”

Proposition 20.1.30. Every finite-dimensional Hermitian space has an orthonormal basis. In fact, every orthonormal list of vectors can be extended to an orthonormal basis.

Proposition 20.1.31. Let $V$ be a Hermitian space with orthonormal basis $b$. Then for all $v, w \in V$,

$$\langle v, w \rangle = [v]_b^* [w]_b. \quad (20.15)$$

Proposition 20.1.32. Let $V$ be a Hermitian space. Every linear form $f : V \to \mathbb{C}$ (Def. 15.1.6) can be written as

$$f(x) = \langle a, x \rangle \quad (20.16)$$

for a unique $a \in V$.

20.2 Hermitian transformations

Definition 20.2.1 (Hermitian linear transformation). Let $V$ be a Hermitian space. Then the linear transformation $\varphi : V \to V$ is Hermitian if for all $v, w \in V$, we have

$$\langle \varphi v, w \rangle = \langle v, \varphi w \rangle. \quad (20.17)$$

Theorem 20.2.2. All eigenvalues of a Hermitian linear transformation are real. ◇
**Proposition 20.2.3.** Let \( \varphi : V \to V \) be a Hermitian linear transformation, and let \( \mathbf{v}_1, \ldots, \mathbf{v}_k \) be eigenvectors of \( \varphi \) with distinct eigenvalues. Then \( \mathbf{v}_i \perp \mathbf{v}_j \) whenever \( i \neq j \).

**Proposition 20.2.4.** Let \( \mathcal{b} \) be an orthonormal basis of the Hermitian space \( V \), and let \( \varphi : V \to V \) be a linear transformation. Then the transformation \( \varphi \) is Hermitian if and only if the matrix \( [\varphi]_\mathcal{b} \) is Hermitian.

Observe that the definition of a Hermitian linear transformation is analogous to that of a symmetric linear transformation of a Euclidean space (Def. 19.4.1), but they are not fully analogous. In particular, while a linear transformation of a Euclidean space that has an orthonormal eigenbasis is necessarily symmetric, the analogous statement in complex spaces involves “normal transformations” (Def. 20.5.1), as opposed to Hermitian transformations.

**Exercise 20.2.5.** Find a linear transformation of \( \mathbb{C}^n \) that has an orthonormal eigenbasis but is not Hermitian.

However, the Spectral Theorem does extend to Hermitian linear transformations.

**Theorem 20.2.6 (Spectral Theorem for Hermitian transformations).** Let \( V \) be a Hermitian space and let \( \varphi : V \to V \) be a Hermitian linear transformation. Then

(a) \( \varphi \) has an orthonormal eigenbasis;

(b) all eigenvalues of \( \varphi \) are real.

**Exercise 20.2.7.** Prove the converse of the Spectral Theorem for Hermitian transformations: If \( \varphi : V \to V \) satisfies (a) and (b), then \( \varphi \) is Hermitian.

In Section 20.6, we shall see a more general form of the Spectral Theorem which extends part (a) to normal transformations (Theorem 20.6.1).

### 20.3 Unitary transformations

In Section 19.3, we introduced orthogonal transformations (Def. 19.3.5), which captured the geometric notion of congruence in Euclidean spaces. The complex analogues of real orthogonal transformations are called unitary transformations.

**Definition 20.3.1 (Unitary transformation).** Let \( V \) be a Hermitian space. Then the transformation \( \varphi : V \to V \) is unitary if it preserves the inner product, i.e.,

\[
\langle \varphi \mathbf{v}, \varphi \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle
\]

(20.18)

for all \( \mathbf{v}, \mathbf{w} \in V \). The set of unitary transformations \( \varphi : V \to V \) is denoted by \( U(V) \).

**Proposition 20.3.2.** The set \( U(V) \) is a group (Def. 14.2.1) under composition.

**Exercise 20.3.3.** The linear transformation \( \varphi : V \to V \) is unitary if and only if \( \varphi \) preserves the norm, i.e., for all \( \mathbf{v} \in \mathbb{C}^n \), we have \( \| \varphi \mathbf{v} \| = \| \mathbf{v} \| \).
Warning. The proof of this is trickier than in the real case (Ex. 19.3.7).

Theorem 20.3.4. Let $\varphi \in U(V)$. Then all eigenvalues of $\varphi$ have absolute value 1. ♦

Proposition 20.3.5 (Consistency of translation). Let $b$ be an orthonormal basis of the Hermitian space $V$, and let $\varphi : V \to V$ be a linear transformation. Then the transformation $\varphi$ is unitary if and only if the matrix $[\varphi]_b$ is unitary.

Theorem 20.3.6 (Spectral Theorem for unitary transformations). Let $V$ be a Hermitian space and let $\varphi : V \to V$ be a unitary transformation. Then

(a) $\varphi$ has an orthonormal eigenbasis;
(b) all eigenvalues of $\varphi$ have unit absolute value.

Exercise 20.3.7. Prove the converse of the Spectral Theorem for unitary transformations: If $\varphi : V \to V$ satisfies (a) and (b) then $\varphi$ is unitary.

20.4 Adjoint transformations in Hermitian spaces

We now study the linear map analogue of the conjugate-transpose of a matrix, known as the Hermitian adjoint.

Theorem 20.4.1. Let $V$ and $W$ be Hermitian spaces, and let $\varphi : V \to W$ be a linear map. Then there exists a unique linear map $\psi : W \to V$ such that for all $v \in V$ and $w \in W$, we have

$$\langle \varphi v, w \rangle = \langle v, \psi w \rangle.$$  (20.19)

Note that the inner product above refers to inner products in two different spaces. To be more specific, we should have written

$$\langle \varphi v, w \rangle_W = \langle v, \psi w \rangle_V.$$  (20.20)

Definition 20.4.2. The linear map $\psi$ whose existence is guaranteed by Theorem 20.4.1 is called the adjoint of $\varphi$ and is denoted $\varphi^*$. So for all $v \in V$ and $w \in W$, we have

$$\langle \varphi v, w \rangle = \langle v, \varphi^* w \rangle.$$  (20.21)

Proposition 20.4.3 (Consistency of translation). Let $V$, $W$, and $\varphi$ be as in the statement of Theorem 20.4.1. Let $b_1$ be an orthonormal basis of $V$ and let $b_2$ be an orthonormal basis of $W$. Then

$$[\varphi^*]_{b_2,b_1} = [\varphi]_{b_1,b_2}^*.$$  (20.22)

TO BE WRITTEN.

20.5 Normal transformations

Definition 20.5.1 (Normal transformation). Let $V$ be a Hermitian space. The transforma-
tion \( \varphi : V \to V \) is *normal* if it commutes with its adjoint, i.e., \( \varphi^* \varphi = \varphi \varphi^* \).

TO BE WRITTEN.

### 20.6 The Complex Spectral Theorem for normal transformations

The main result of this section is the generalization of the Spectral Theorem to normal transformations.

**Theorem 20.6.1** (Complex Spectral Theorem). *Let \( \varphi : V \to V \) be a linear transformation. Then \( \varphi \) has an orthonormal eigenbasis if and only if \( \varphi \) is normal.* \( \diamond \)
Chapter 21

(\mathbb{R}, \mathbb{C}) The Singular Value Decomposition

In this chapter, we discuss matrices over \mathbb{C}, but every statement of this chapter holds over \mathbb{R} as well, if we replace matrix adjoints by transposes and the word “unitary” by “orthogonal.”

21.1 The Singular Value Decomposition

In this chapter we study the “Singular Value Decomposition,” an important tool in many areas of math and computer science.

Notation 21.1.1. For \( r \leq \min\{k, n\} \), the matrix \( \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_r) \) is the \( k \times n \) matrix with \( \sigma_i \) (\( 1 \leq i \leq r \)) in the \((i, i)\) entry and 0 everywhere else, i.e., the matrix

\[
\begin{pmatrix}
\sigma_1 & 0 & & \\
0 & \sigma_2 & & \\
& \ddots & \ddots & \\
0 & & \ddots & \sigma_r \\
0 & & & 0 \\
\end{pmatrix}
\]

Note that such a matrix is a “diagonal” matrix which is not necessarily square.

Theorem 21.1.2 (Singular Value Decomposition). Let \( A \in \mathbb{C}^{k \times n} \). Then there exist unitary matrices \( S \in U(k) \) and \( T \in U(n) \) such that

\[
S^* A T = \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_r)
\]

(21.1)

where each “singular value” \( \sigma_i \) is real, \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \), and \( r = \text{rk} A \). If \( A \in \mathbb{R}^{k \times n} \), then we can let \( S \) and \( T \) be real, i.e., \( S \in O(k) \) and \( T \in O(n) \).

Theorem 21.1.3. The singular values of \( A \) are the square roots of the nonzero eigenvalues of \( A^* A \).

The remainder of this section is dedicated to proving the following restatement of Theorem 21.1.2.

Theorem 21.1.4 (Singular Value Decomposition, restated). Let \( V \) and \( W \) be Hermitian spaces with \( \dim V = n \) and \( \dim W = k \). Let \( \varphi : V \to W \) be a linear map of rank \( r \). Then there exist orthonormal bases \( e \) and \( f \) of \( V \) and
CHAPTER 21. \((\mathbb{R}, \mathbb{C})\) THE SINGULAR VALUE DECOMPOSITION

\(W, \text{ respectively, and real numbers } \sigma_1 \geq \cdots \geq \sigma_r > 0 \text{ such that } \varphi e_i = \sigma_i f_i \text{ and } \varphi^* f_i = \sigma_i e_i \text{ for } i = 1, \ldots, r, \text{ and } \varphi e_j = 0 = \varphi^* f_j \text{ for } j > r. \)

Exercise 21.1.5. Show that Theorem 21.1.4 is equivalent to Theorem 21.1.2.

Exercise 21.1.6. Let \(V\) and \(W\) be Hermitian spaces and let \(\varphi : V \to W\) be a linear map of rank \(r\). Let \(e = (e_1, \ldots, e_r, \ldots, e_n)\) be an orthonormal eigenbasis of \(\varphi^* \varphi\) (why does such a basis exist?) with corresponding eigenvalues \(\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n\). Let \(\sigma_i = \sqrt{\lambda_i}\) for \(u = 1, \ldots, r, \) Let \(f_i = \frac{1}{\sigma_i} e_i\) for \(i = 1, \ldots, r, \) Then for \(1 \leq i \leq r, \)

(a) \(\varphi e_i = \sigma_i f_i;\)

(b) \(\varphi^* f_i = \sigma_i e_i.\)

Exercise 21.1.7. The \(f_i\) of the preceding exercise are orthonormal.


21.2 Low-rank approximation

In this section, we use the Singular Value Decomposition to find low-rank approximations to matrices, that is, the matrix of a given rank which is “closest” to a specified matrix under the operator norm (\(\|\|\) Def. 13.1.1).

Definition 21.2.1 (Truncated matrix). Let \(D = \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_r)\) be a rank-\(r\) matrix (so \(\sigma_1, \ldots, \sigma_r \neq 0\)). The rank-\(\ell\) truncation (\(\ell \leq r\)) of \(D\), denoted \(D_\ell\), is the \(k \times n\) matrix \(D_\ell = \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_\ell)\).

The next theorem explains how the Singular Value Decomposition helps us find low-rank approximations to matrices.

Theorem 21.2.2 (Nearest low-rank matrix). Let \(A \in \mathbb{C}^{k \times n}\) be a matrix of rank \(r\) with singular values \(\sigma_1 \geq \cdots \geq \sigma_r > 0\). Define \(S, T, \) and \(D\) as guaranteed by the Singular Value Decomposition Theorem so that \(S^* AT = \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_r) = D.\) (21.2)

Given \(\ell \leq r,\) let

\[D_\ell = \text{diag}_{k \times n}(\sigma_1, \ldots, \sigma_\ell)\]  (21.3)

and define \(B_\ell = SD_\ell T^*.\) Then \(B_\ell\) is the matrix of rank at most \(\ell\) which is nearest to \(A\) under the operator norm, i.e., \(\text{rk} B_\ell = \ell\) and for all \(B \in \mathbb{C}^{k \times n}, \) if \(\text{rk} B \leq \ell, \) then \(\|A - B_\ell\| \leq \|A - B\|.\) \(\textcircled{\text{0}}\)

Exercise 21.2.3. Let \(A, B_\ell, D,\) and \(D_\ell\) be as in the statement of Theorem 21.2.2. Show

(a) \(\|A - B_\ell\| = \|D - D_\ell\|;\)

(b) \(\|D - D_\ell\| = \sigma_{\ell+1}.\)

As with the proof of the Singular Value Decomposition Theorem, Theorem 21.2.2 is easier to prove in terms of linear maps. We restate the theorem as follows.
21.2. LOW-RANK APPROXIMATION

**Theorem 21.2.4.** Let $V$ and $W$ be Hermitian spaces with orthonormal bases $e$ and $f$, respectively, and let $\varphi : V \to W$ be a linear map such that $\varphi e_i = \sigma_i f_i$ for $i = 1, \ldots, r$. Define the truncated map $\varphi_\ell : V \to V$ by

$$
\varphi_\ell e_i = \begin{cases} 
\sigma_i f_i & 1 \leq i \leq \ell \\
0 & \text{otherwise}
\end{cases} \quad (21.4)
$$

Then whenever $\psi : V \to W$ is a linear map of rank $\leq \ell$, we have $\|\varphi - \varphi_\ell\| \leq \|\varphi - \psi\|$. ◊

**Exercise 21.2.5.** Show that Theorem 21.2.4 is equivalent to Theorem 21.2.2.

**Exercise 21.2.6.** Let $\varphi$ and $\varphi_\ell$ be as in the statement of the preceding theorem. Show $\|\varphi - \varphi_\ell\| = \sigma_{\ell+1}$.

It follows that in order to prove Theorem 21.2.4, it suffices to show that for all linear maps $\psi : V \to W$ of rank $\leq \ell$, we have $\|\varphi - \psi\| \geq \sigma_{\ell+1}$.

**Exercise 21.2.7.** Let $\psi : V \to W$ be a linear map of rank $\leq \ell$. Show that there exists $v \in \ker \psi$ such that

$$
\frac{\| (\varphi - \psi)v \|}{\|v\|} \geq \sigma_{\ell+1} . \quad (21.5)
$$

**Exercise 21.2.8.** Complete the proof of Theorem 21.2.4 hence of Theorem 21.2.2.

**Exercise 21.2.9.** Show that the matrix $B_\ell$ whose existence is guaranteed by Theorem 21.2.2 is unique, i.e., if there is a matrix $B \in \mathbb{C}^{k \times \ell}$ of rank $\ell$ such that, $\|A - B'\| \leq \|A - B\|$ for all rank-$\ell$ matrices $B' \in \mathbb{C}^{k \times \ell}$, then $B = B_\ell$.

In fact, the rank-$\ell$ matrix guaranteed by Theorem 21.2.2 to be nearest to $A$ under the operator norm is also the rank-$\ell$ matrix nearest to $A$ under the Frobenius norm (Theorem 13.2.1).

**Theorem 21.2.10.** The statement of Theorem 21.2.2 holds for the same matrix $B_\ell$ when the operator norm is replaced by the Frobenius norm. That is, we also have $\|A - B_\ell\|_F \leq \|A - B\|_F$ for all rank-$\ell$ matrices $B \in \mathbb{C}^{k \times n}$. ◊
Chapter 22

(R) Finite Markov Chains

22.1 Stochastic matrices

Definition 22.1.1 (Nonnegative matrix). We say that $A$ is a nonnegative matrix if all entries of $A$ are nonnegative.

Definition 22.1.2 (Stochastic matrix). A square matrix $A = (\alpha_{ij}) \in M_n(\mathbb{R})$ is stochastic if it is nonnegative and the rows of $A$ sum to 1, i.e., $\sum_{j=1}^{n} \alpha_{ij} = 1$ for all $i$.

Examples 22.1.3. The following matrices are stochastic.

(a) $\frac{1}{n} J_n$

(b) Permutation matrices

(c) \[
\begin{pmatrix}
0.8 & 0.2 \\
0.3 & 0.7
\end{pmatrix}
\]

Exercise 22.1.4. Let $A \in \mathbb{R}^{k \times n}$ and $B \in \mathbb{R}^{n \times m}$. Prove that if $A$ and $B$ are stochastic matrices then $AB$ is a stochastic matrix.

Exercise 22.1.5.

(a) Let $A$ be a stochastic matrix. Prove that $A1 = 1$.

(b) Show that the converse is false.

(c) Show that $A$ is stochastic if and only if $A$ is nonnegative and $A1 = 1$.

Definition 22.1.6 (Probability distribution). A probability distribution is a list of nonnegative numbers which add to 1.

Fact 22.1.7. Every row of a stochastic matrix is a probability distribution.

22.2 Finite Markov Chains

A finite Markov Chain is a stochastic process defined by a finite number of states and constant transition probabilities. We consider the trajectory of a particle that can move between states at discrete time steps with the given transition probabilities. In this section, we denote by $\Omega = [n]$ the finite set of states.

FIGURE HERE

This is described more formally below.
Definition 22.2.1 (Finite Markov Chain). Let \( \Omega = [n] \) be a set of states. We denote by \( X_t \in \Omega \) the position of the particle at time \( t \). The transition probability \( p_{ij} \) is the probability that the particle moves to state \( j \) at time \( t + 1 \), given that it is in state \( i \) at time \( t \), i.e.,

\[
p_{ij} = P(X_{t+1} = j \mid X_t = i) .
\]

(22.1)

In particular, each \( p_{ij} \geq 0 \) and \( \sum_{j=1}^{n} p_{ij} = 1 \) for every \( i \). The infinite sequence \((X_0, X_1, X_2, \ldots)\) is a stochastic process.

Definition 22.2.2 (Transition matrix). The transition matrix corresponding to our finite Markov Chain is the matrix \( T = (p_{ij}) \).

Fact 22.2.3. The transition matrix \( T \) of a finite Markov Chain is a stochastic matrix, and every stochastic matrix is the transition matrix of a finite Markov Chain.

Notation 22.2.4 (r-step transition probability). The \( r \)-step transition probability from state \( i \) to state \( j \) is denoted \( p_{ij}^{(r)} \) and is defined by

\[
p_{ij}^{(r)} = P(X_{t+r} = j \mid X_t = i) .
\]

(22.2)

Exercise 22.2.5 (Evolution of Markov Chains, I). Let \( T = (p_{ij}) \) be the transition matrix corresponding to a finite Markov Chain. Show that \( T^r = \left( p_{ij}^{(r)} \right) \) where

Definition 22.2.6. Let \( q_{t,i} \) be the probability that the particle is in state \( i \) at time \( t \). We define \( q_t = (q_{t,1}, \ldots, q_{t,n}) \) to be the distribution of the particle at time \( t \).

Fact 22.2.7. For every \( t \geq 0 \), \( \sum_{i=1}^{n} q_{t,i} = 1 \).

Exercise 22.2.8 (Evolution of Markov Chains, II). Let \( T \) be the transition matrix of a finite Markov Chain. Show that \( q_{t+1} = q_t T \) and conclude that \( q_t = q_0 T^t \).

Definition 22.2.9 (Stationary distribution). The probability distribution \( q \) is a stationary distribution if \( q = q T \), i.e., if \( q \) is a left eigenvector (\( \mathbb{R} \text{ Def. 8.1.15} \)) with eigenvalue 1.

Proposition 22.2.10. Let \( A \in M_n(\mathbb{R}) \) be a stochastic matrix. Show that \( A \) has a left eigenvector with eigenvalue 1.

The preceding proposition in conjunction with the next theorem shows that every stochastic matrix has a stationary distribution.

Recall (\( \mathbb{R} \text{ Def. 22.1.1} \)) that a nonnegative matrix is a matrix whose entries are all nonnegative; a nonnegative vector is defined similarly. Do not prove the following theorem.

Theorem 22.2.11 (Perron-Frobenius). Every nonnegative square matrix has a nonnegative eigenvector.

Corollary 22.2.12. Every finite Markov Chain has a stationary distribution.

Proposition 22.2.13. If \( T \) is the transition matrix of a finite Markov Chain and \( \lim_{r \to \infty} T^r = L \) exists, then every row of \( L \) is a stationary distribution.
In order to determine which Markov Chains have transition matrices that converge, we study the directed graphs associated with finite Markov Chains.

### 22.3 Digraphs
TO BE WRITTEN.

### 22.4 Digraphs and Markov Chains

Every finite Markov Chain has an associated digraph, where $i \rightarrow j$ if and only if $p_{ij} \neq 0$.

**FIGURE HERE**

**Definition 22.4.1 (Irreducible Markov Chain).** A finite Markov Chain is *irreducible* if its associated digraph is strongly connected.

**Proposition 22.4.2.** If $T$ is the transition matrix of an irreducible finite Markov Chain and $\lim_{r \to \infty} T^r = L$ exists, then all rows of $L$ are the same, so $\text{rk } L = 1$.

**Definition 22.4.3 (Ergodic Markov Chain).** A finite Markov Chain is *ergodic* if its associated digraph is strongly connected and aperiodic.

The following theorem establishes a sufficient condition under which the probability distribution of a finite Markov Chain converges to the stationary distribution. For irreducible Markov Chains, this is necessary and sufficient.

**Theorem 22.4.4.** If $T$ is the transition matrix of an ergodic finite Markov Chain, then $\lim_{r \to \infty} T^r$ exists.

**Proposition 22.4.5.** For irreducible Markov Chains, the stationary distribution is unique.

### 22.5 Finite Markov Chains and undirected graphs
TO BE WRITTEN.

### 22.6 Additional exercises

**Definition 22.6.1 (Doubly stochastic matrix).** A matrix $A = (\alpha_{ij})$ is *doubly stochastic* if both $A$ and $A^T$ are stochastic, i.e., $0 \leq \alpha_{ij} \leq 1$ for all $i, j$ and every row and column sum is equal to 1.

**Examples 22.6.2.** The following matrices are doubly stochastic.

(a) $\frac{1}{n} J_n$
(b) Permutation matrices
(c) \[
\begin{pmatrix}
0.1 & 0.3 & 0.6 \\
0.2 & 0.5 & 0.3 \\
0.7 & 0.2 & 0.1
\end{pmatrix}
\]

**Theorem* 22.6.3 (Birkhoff’s Theorem).** The set of doubly stochastic matrices is the convex hull (\ref{Def:ConvexHull} Def. 5.3.6) of the set of permutation
matrices (Def. 2.4.3). In other words, every doubly stochastic matrix can be expressed as a convex combination (Def. 5.3.1) of permutation matrices.
Chapter 23

More Chapters

TO BE WRITTEN.
Chapter 24

Hints

24.1 Column Vectors

1.1.17 Exercise Solution
(b) Consider the sums of the entries in each column vector.

1.2.7 Exercise Solution
Show that the sum of two vectors of zero weight has zero weight and that scaling a zero-weight vector produces a zero-weight vector.

1.2.8 Exercise Solution
What are the subspaces of $\mathbb{R}^n$ which are spanned by one vector?

1.2.9 Exercise Solution
(c) The “if” direction is trivial. For the “only if” direction, begin with vectors $w_1 \in W_1 \setminus W_2$ and $w_2 \in W_2 \setminus W_1$. Where is $w_1 + w_2$?

1.2.12 Exercise Solution
Prove (a) and (b) together. Prove that the subspace defined by (b) satisfies the definition.

1.2.15 Exercise Solution
Preceding exercise.

1.2.16 Exercise Solution
Show that $\text{span}(T) \subseteq \text{span}(S)$.

1.2.18 Exercise Solution
It is clear that $U_1 + U_2 \subseteq \text{span}(U_1 \cup U_2)$. For the reverse inclusion you need to show that a linear combination from $U_1 \cup U_2$ can always be written as a sum $u_1 + u_2$ for some $u_1 \in U_1$ and $u_2 \in U_2$.

1.3.9 Exercise Solution
You need to find $\alpha_1, \alpha_2, \alpha_3$ such that
$$\begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} -2 \\ -8 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$ This gives you a system of linear equations. Find a nonzero solution.

1.3.11 Exercise Solution
Assume $\sum \alpha_i v_i = 0$. What condition on $\alpha_j$ allows you to express $v_j$ in terms of the other vectors?

1.3.13 Exercise Solution
What does the empty sum evaluate to?

1.3.15 Exercise Solution
Find a nontrivial linear combination of \( \mathbf{v} \) and \( \mathbf{v} \) that evaluates to 0.

1.3.16 Exercise Solution
The only linear combinations of the list \( \mathbf{v} \) are of the form \( \alpha \mathbf{v} \) for \( \alpha \in \mathbb{F} \).

1.3.17 Exercise Solution
Ex. 1.3.13

1.3.21 Exercise Solution
If \( \sum \alpha_i \mathbf{v}_i = 0 \) and not all the \( \alpha_i \) are 0, then it must be the case that \( \alpha_{k+1} \neq 0 \) (why?).

1.3.22 Exercise Solution
Prop. 1.3.11

1.3.23 Exercise Solution
What is the simplest nontrivial linear combination of such a list which evaluates to zero?

1.3.24 Exercise Solution
Combine Prop. 1.3.15 and Fact 1.3.20

1.3.27 Exercise Solution
Begin with a basis. How can you add one more vector that will satisfy the condition?

1.3.41 Exercise Solution
Suppose not. Use Lemma 1.3.21 to show that \( \text{span}(\mathbf{w}_1, \ldots, \mathbf{w}_m) \leq \text{span}(\mathbf{v}_2, \ldots, \mathbf{v}_k) \) and arrive at a contradiction.

1.3.42 Exercise Solution
Use the Steinitz exchange lemma to replace \( \mathbf{v}_1 \) by some \( \mathbf{w}_{i_1} \), then \( \mathbf{v}_2 \) by some \( \mathbf{w}_{i_2} \), etc. So in the end, \( \mathbf{w}_{i_1}, \ldots, \mathbf{w}_{i_k} \) are linearly independent.

1.3.45 Exercise Solution
Use the standard basis of \( \mathbb{F}^k \) and the preceding exercise.

1.3.46 Exercise Solution
It is clear that \( \text{rk}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \leq \dim(\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)) \) (why?). For the other direction, first assume that the \( \mathbf{v}_i \) are linearly independent, and show that \( \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \) cannot contain a list of more than \( k \) linearly independent vectors.

1.3.51 Exercise Solution
Show that if this list were not linearly independent, then \( U_1 \cap U_2 \) would contain a nonzero vector.

1.3.52 Exercise Solution
Ex. 1.3.51

1.3.53 Exercise Solution
Start with a basis of \( U_1 \cap U_2 \).

1.4.9 Exercise Solution
Only the zero vector. Why?

1.4.10 Exercise Solution
Express \( 1 \cdot \mathbf{x} \) in terms of the entries of \( \mathbf{x} \).

1.5.4 Exercise Solution
Begin with a linear combination \( W \) of the vectors in \( S \) which evaluates to zero. Conclude
that all coefficients must be 0 by taking the dot product of \( W \) with each member of \( S \).

1.6.4 Exercise Solution
Incidence vectors.

24.2 Matrices

2.2.3 Exercise Solution
Prove only one of these, then infer the other using the transpose (Ex. 2.2.2).

2.2.17 Exercise Solution
(d3) The size is the sum of the squares of the entries. This is the Frobenius norm of the matrix.

2.2.19 Exercise Solution
Show, by induction on \( k \), that the \((i, j)\) entry of \( A^k \) is zero whenever \( j \leq i + k - 1 \).

2.2.27 Exercise Solution
Write \( B \) as a sum of matrices of the form 
\[
\begin{bmatrix}
0 & \cdots & 0 & b_i & 0 & \cdots & 0
\end{bmatrix}
\]
and then use distributivity.

2.2.28 Exercise Solution
The \( i \)-th entry of \( Ax \) is the dot product of the \( i \)-th row of \( A \) with \( x \).

2.2.29 Exercise Solution
Apply the preceding exercise to \( B = I \).

2.2.30 Exercise Solution
Consider \( (A - B)x \).

2.2.31 Exercise Solution
Prop. 2.2.28

2.2.32 Exercise Solution
Prop. 2.2.29

2.2.33 Exercise Solution
Prop. 2.2.29

2.2.38 Exercise Solution
What is the \((i, i)\) entry of \( AB \)?

2.2.39 Exercise Solution
Preceding exercise.

2.2.41 Exercise Solution
If \( \text{Tr}(AB) = 0 \) for all \( A \), then \( B = 0 \).

2.3.2 Exercise Solution
Induction on \( k \). Use the preceding exercise.

2.3.4 Exercise Solution
Induction on the degree of \( f \).

2.3.6 Exercise Solution
Observe that the off-diagonal entries do not affect the diagonal entries.

2.3.7 Exercise Solution
Induction on \( k \). Use the preceding exercise.

2.5.1 Exercise Solution
Ex. 2.2.2

2.5.5 Exercise Solution
2.5.6 Exercise Solution
Multiply by the matrix with a 1 in the \((i,j)\) position and 0 everywhere else.

2.5.7 Exercise Solution
(a) Trace.

2.5.8 Exercise Solution
Incidence vectors (Def. 1.6.2).

2.5.13 Exercise Solution
(b) That particular circulant matrix is the one corresponding to the cyclic permutation of all elements, i.e., the matrix \(C = C(0,1,0,0,\ldots,0)\).

24.3 Matrix Rank

3.1.5 Exercise Solution
Consider the matrix \([A \mid B]\) (the \(k \times 2n\) matrix obtained by concatenating the columns of \(B\) onto \(A\)), and compare its column rank to the quantities above.

3.3.2 Exercise Solution
Elementary column operations do not change the column space of a matrix.

3.3.12 Exercise Solution
Consider the \(k \times r\) submatrix formed by taking \(r\) linearly independent columns.

3.5.9 Exercise Solution
Extend a basis of \(U\) to a basis of \(\mathbb{F}^n\), and consider the action of the matrix \(A\) on this basis.

3.6.7 Exercise Solution
Ex. 2.2.19.

24.4 Theory of Systems of Linear Equations I: Qualitative Theory

24.5 Affine and Convex Combinations

5.1.14 Exercise Solution
(b) Let \(v_0 \in S\) and let \(U = \{s - v_0 \mid s \in S\}\). Show that

(i) \(U \leq \mathbb{F}^n\);

(ii) \(S\) is a translate of \(U\).

For uniqueness, you need to show that if \(U + u = V + v = S\), then \(U = V\).
24.6  The Determinant

6.2.22  Exercise  Solution
Identity matrix.

6.2.23  Exercise  Solution
Consider the $n!$ term expansion of the determinant (Eq. (6.20)). What happens with the contributions of $\sigma$ and $\sigma^{-1}$? What if $\sigma = \sigma^{-1}$? What if $\sigma$ has a fixed point?

6.3.7  Exercise  Solution
Let $f_{e_n}$ denote the number of fixed-point-free even permutations of the set $\{1, \ldots, n\}$, and let $f_{o_n}$ denote the number of fixed-point-free odd permutations of the set $[n]$. Write $f_{o_n} - f_{e_n}$ as the determinant of a familiar matrix.

6.4.6  Exercise  Solution
$\det A = \pm 1$.

6.4.7  Exercise  Solution
Ex. 6.2.23

24.7  Theory of Systems of Linear Equations II: Cramer’s Rule

24.8  Eigenvectors and Eigenvalues

8.1.22  Exercise  Solution

Use Rank-Nullity.

8.5.6  Exercise  Solution
Prop. 2.3.4

24.9  Orthogonal Matrices

24.10  The Spectral Theorem

24.11  Bilinear and Quadratic Forms

11.1.5  Exercise  Solution
Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$ (Def. 1.3.34). Let $\alpha_i = f(e_i)$. Show that $a = (\alpha_1, \ldots, \alpha_n)^T$ works.

11.1.10  Exercise  Solution
Generalize the hint to Theorem 11.1.5

11.1.13  Exercise  Solution
(c) Skew symmetric matrices. Ex. 6.2.21

11.3.3  Exercise  Solution
Give a very simple expression for $B$ in terms of $A$.

11.3.9  Exercise  Solution
Use the fact that $\lambda$ is real.

11.3.15  Exercise  Solution
Interlacing.

11.3.23 Exercise Solution
(b) Triangular matrices.

24.12 Complex Matrices

12.2.8 Exercise Solution

(a)

(b) Prove that your example must have rank 1.

12.4.9 Exercise Solution
Induction on $n$.

12.4.15 Exercise Solution
Prop. 12.4.6

12.4.19 Exercise Solution
Theorem 12.4.18

12.5.4 Exercise Solution
Pick $\lambda_0, \ldots, \lambda_{n-1}$ to be complex numbers with unit absolute value. Let $w = (\lambda_0, \ldots, \lambda_{n-1})^T$. Prove that the entries of $\frac{1}{\sqrt{n}} F w$ generate a unitary circulant matrix, and that all unitary circulant matrices are of this form. The $\lambda_i$ are the eigenvalues of the resulting circulant matrix.

24.13 Matrix Norms

13.1.2 Exercise Solution
Show that $\max_{x \in \mathbb{R}^n} \frac{\|Ax\|}{\|x\|}$ exists. Why does this suffice?

13.1.13 Exercise Solution
Use the fact that $f_{AB} = f_{BA} \cdot t^{n-k}$ (Prop. 8.4.16).

24.14 Preliminaries

24.15 Vector Spaces

15.3.14 Exercise Solution
Use the fact that a polynomial of degree $n$ has at most $n$ roots.

15.4.9 Exercise Solution
Ex. 14.4.57

15.4.10 Exercise Solution
Ex. 16.4.10

15.4.13 Exercise Solution
Start with a basis of $U_1 \cap U_2$.

15.4.14 Exercise Solution
It is an immediate result of Cor. 15.4.3 that
every list of \( k + 1 \) vectors in \( \mathbb{R}^k \) is linearly dependent.

15.4.18 Exercise Solution
Their span is not changed.

24.16 Linear Maps

16.2.6 Exercise Solution
Coordinate vector.

16.3.7 Exercise Solution
Let \( e_1, \ldots, e_k \) be a basis of \( \ker(\varphi) \). Extend this to a basis \( e_1, \ldots, e_n \) of \( V \), and show that \( \varphi(e_{k+1}), \ldots, \varphi(e_n) \) is a basis for \( \text{im}(\varphi) \).

16.4.32 Exercise Solution

(a)

(b)

(c) Prop. 16.4.14

(d)

16.6.7 Exercise Solution
Write \( A = [\varphi]_{\text{old}} \) and \( A' = [\varphi]_{\text{new}} \). Show that for all \( x \in \mathbb{F}^n \) (\( n = \dim V \)), we have \( A'x = T^{-1}A'Sx \).

24.17 Block Matrices

24.18 Minimal Polynomials of Matrices and Linear Transformations

24.19 Euclidean Spaces

19.1.7 Exercise Solution
Derive this from the Cauchy-Schwarz inequality.

19.1.21 Exercise Solution
Theorem 11.1.5

19.3.20 Exercise Solution
Prop. 19.1.21

19.4.19 Exercise Solution
\( f'(0) = 0 \).

19.4.22 Exercise Solution
Define \( U = \{v_0\}^\perp \), and for each \( u \in U \setminus \{0\} \), consider the function \( f_u : \mathbb{R} \rightarrow \mathbb{R} \) defined by \( f_u(t) = R_\varphi(v_0 + tu) \). Apply the preceding exercise to this function.
24.20 Hermitian Spaces

20.1.19 Exercise Solution
Derive this from the Cauchy-Schwarz inequality.

20.1.32 Exercise Solution
Theorem 11.1.5.

20.2.2 Exercise Solution
For an eigenvalue \( \lambda \), you need to show that \( \lambda = \bar{\lambda} \). The proof is just one line. Consider the expression \( x^*Ax \).

20.2.5 Exercise Solution
Which diagonal matrices are Hermitian?

20.4.1 Exercise Solution
Prop. 20.1.32.

24.21 The Singular Value Decomposition

21.2.7 Exercise Solution
It suffices to show that there exists \( v \in \ker \psi \) such that \( ||\varphi v|| \geq \sigma_{\ell+1}||v|| \) (why?). Pick \( v \in \ker \psi \cap \text{span}(e_1, \ldots, e_{\ell+1}) \) and show that this works.

24.22 Finite Markov Chains

22.1.5 Exercise Solution
For a general matrix \( B \in M_n(\mathbb{R}) \), what is the \( i \)-th entry of \( B1 \)?

22.2.10 Exercise Solution
Prop. 8.4.15.

22.6.3 Exercise Solution
Marriage Theorem.
Chapter 25

Solutions

25.1 Column Vectors

1.1.16 Exercise

\[
\begin{pmatrix}
-5 \\
-1 \\
11
\end{pmatrix}
= 2 \begin{pmatrix}
-2 \\
1 \\
7
\end{pmatrix}
- \frac{1}{2} \begin{pmatrix}
2 \\
6 \\
6
\end{pmatrix}
\]

25.2 Matrices

25.3 Matrix Rank

25.4 Theory of Systems of Linear Equations I: Qualitative Theory

25.5 Affine and Convex Combinations

25.6 The Determinant

25.7 Theory of Systems of Linear Equations II: Cramer’s Rule

25.8 Eigenvectors and Eigenvalues

8.1.3 Exercise

Every vector in \( \mathbb{R}^n \) is an eigenvector of \( I_n \) with eigenvalue 1.

8.1.22 Exercise

\[
\dim U_\lambda = n - \text{rk}(\lambda I - A).
\]
25.9 Orthogonal Matrices

25.10 The Spectral Theorem

10.2.2 Exercise Hint

\[ v^T A v = \left( \sum_{i=1}^{n} \alpha_i b_i^T \right) A \left( \sum_{i=1}^{n} \alpha_i b_i \right) \]

\[ = \left( \sum_{i=1}^{n} \alpha_i b_i^T \right) \left( \sum_{i=1}^{n} \alpha_i \lambda_i b_i \right) \]

\[ = \left( \sum_{i=1}^{n} \alpha_i b_i^T \right) \left( \sum_{i=1}^{n} \alpha_i \lambda_i b_i \right) \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \alpha_i \alpha_j b_j^T b_i \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \alpha_i \alpha_j (b_j \cdot b_i) \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \alpha_i \alpha_j \delta_{ij} \]

\[ = \sum_{i=1}^{n} \lambda_i \alpha_i^2 \]

Hence \( \lambda > 0 \).

Now suppose \( A \) is symmetric and all of its eigenvalues are positive. By the Spectral Theorem, \( A \) has an orthonormal eigenbasis, say \( b = (b_1, \ldots, b_n) \), with \( Ab_i = \lambda_i b_i \) for all \( i \). Let \( v \in \mathbb{F}^k \) be a nonzero vector. Then we can write \( v = \sum \alpha_i b_i \) for some scalars \( \alpha_i \). Then

\[ v^T A v = \sum_{i=1}^{n} \lambda_i \alpha_i^2 > 0 \]

so \( A \) is positive definite.

25.11 Bilinear and Quadratic Forms

10.2.4 Exercise Hint

First suppose \( A \) is positive definite. Let \( v \) be an eigenvector of \( A \). Then, because \( v \neq 0 \), we have

\[ 0 < v^T A v = v^T \lambda v = \lambda v^T v = \lambda ||v||^2 \]

11.3.3 Exercise Hint

\[ B = \frac{A + A^T}{2} \] Verify that this and only this matrix works.
25.12 Complex Matrices

25.13 Matrix Norms

25.14 Preliminaries

25.15 Vector Spaces

25.16 Linear Maps

25.17 Block Matrices

25.18 Minimal Polynomials of Matrices and Linear Transformations

\[ m_A = \prod_{i=1}^{n} (t - \lambda_i) \] where \( \prod' \) means that we take each factor only once (so eigenvalues \( \lambda_i \) with multiplicity greater than 1 only contribute one factor of \( t - \lambda_i \)).

25.19 Euclidean Spaces

25.20 Hermitian Spaces

25.21 The Singular Value Decomposition

25.22 Finite Markov Chains
Index

0, 410, 13
1, 413, 20

Affine basis, 43, 43
Affine combination, 41, 41, 44
Affine hull, 41, 41, 42, 45
Affine independence, 43, 43
Affine subspace, 39
of $\mathbb{F}^k$, 41, 41, 44
codimension of, 43
dimension of, 43, 43
intersection of, 41, 42

All-ones matrix, see $J$ matrix
All-ones vector, see 1

Basis
of a subspace of $\mathbb{F}^k$, 10, 12
of a subspace of $\mathbb{F}^k$, 11
of $\mathbb{F}^k$, 11, 40

Bilinear form
nonsingular, 73
over $\mathbb{F}^n$, 68, 68, 69
nonsingular, 69, 69
representation theorem for, 69

$\mathbb{C}^n$
orthonormal basis of, 77

Cauchy-Binet formula, 55

Cayley-Hamilton Theorem, 62, 62, 63, 78
for diagonal matrices, 62
for diagonalizable matrices, 63

Characteristic polynomial, 60, 60, 62
of nilpotent matrices, 61
of similar matrices, 63

Codimension, 34
of an affine subspace, 43

Cofactor, 53

Cofactor expansion, 53, 53, 54
skew, 54

Column space, 28, 32

Column vector
orthogonality of, 76

addition of, 4, 4
dot product of, 12, 12, 14, 20, 64, 69, 74
isotropic, 74, 74
linear combination of, 8, 8, 6, 12, 20
linear dependence of, 8, 8, 52
Linear independence of, 12
linear independence of, 8, 8, 10, 13, 14
26, 32, 35, 39, 52, 57
multiplication by a scalar, 3, 3
norm of, 76
orthogonality, 13, 13, 14
orthogonality of, 20
with respect to a bilinear form, 73
parallel, 10, 10
Column-rank, 28, 28, 31
full, 28, 33
Column-space, 28
Companion matrix, 61, 61, 62
Complex conjugate, 75
Complex number, 75, 75
absolute value of, 75
imaginary part of, 75
magnitude of, 75
norm of, 75
real part of, 75
unit norm, 75, 75
Convex combination, 44
Convex hull, 44, 44, 45
Convex set, 44, 44, 46
dimension of, 45
intersection of, 44
Coordinates, 66
Corank
of a matrix, 35, 35
of a set, 34
Courant-Fischer Theorem, 67
Cramer’s Rule, 56, 56
Determinant, 40, 48, 50, 50, 53, 61, 72
corner, 72, 72
multilinearity of, 51
of a matrix product, 51
Diagonalizable matrix, 59, 59
Dimension, 10, 11, 12

of a convex set, 43
of an affine subspace, 43, 43
Discrete Fourier Transform, 79, 79
Dot product, 13, 12, 14, 20, 64, 69, 74
Dual space
of $\mathbb{F}^n$, 68

Eigenbasis
of $\mathbb{F}^n$, 57, 57, 58
orthonormal, 65, 66, 78
Eigenspace, 58, 58
Eigenvalue
interlacing of, 67
of a matrix, 57, 57, 58, 60, 61, 71, 72, 80
81
algebraic multiplicity of, 61, 61
geometric multiplicity of, 58, 58, 61
left, 58, 58, 61
of a real symmetric matrix, 66
Eigenvector
of a matrix
left, 58
of a matrix, 57, 57, 58
left, 58
Elementary column operation, 29, 29, 31, 35
52
and the determinant, 51
Elementary matrices, 30
Elementary matrix, 30
Elementary row operation, 29, 29, 32, 35
Empty list, 6
Empty sum, 4
Euclidean norm, 13, 52, 64, 77
complex, 76
$\mathbb{F}^k$, 4, 15
Matrix, 15
addition of, 17
adjugate of, 34
augmented, 39
block, 118
  multiplication of, 119
block-diagonal, 118
block-triangular, 118
characteristic polynomial of, 60
  60 62
circulant, 26
  26 79
commutator of, 25
conjugate-transpose of, 75
  75 76
corank of, 35
corner, 72
definiteness of, 71
  71 72
diagonal, 16
  16 19 22 25 57 71 77
determinant of, 50
  59 59 78
doubly stochastic, 146
  146 146 147
exponentiation, 16
Hermitian, 77
  77 78
Hermitian adjoint of, 75
identity, see Identity matrix
  indefinite, 71
  integral, 15
  inverse, 34
  34 34 34
  34 35
  35
  kernel of, 34
  left inverse of, 33
  lower triangular, 16
  multiplication, 18
  and the determinant, 51
  associativity of, 18
  distributivity of, 18
  multiplication by a scalar, 17
  negative definite, 71
  negative semidefinite, 71
  nilpotent, 20
  20 58 59
  characteristic polynomial of, 61
nonnegative, 144
nonsingular, 32
  32 32 33 36 39 54 56 58
  69
orthogonal, 64
  64 65 80 81
orthogonally similar, 64
  64 66
permutation, see Permutation matrix
positive definite, 66
  67 71 80
positive semidefinite, 71
rank, 32
  32 56 76
right inverse of, 33
  33 34 40
rotation, see Rotation matrix
scalar, 19
self-adjoint, 77
  77 78
similarity of, 58
  58 59 62
skew-symmetric, 52
square, 15
stochastic, 80
  80 144 144
strictly lower triangular, 16
strictly upper triangular, 16
symmetric, 17
  17 25 64 80 81
trace of, 21
transpose, 17
  17 18 25
determinant of, 50
unitary, 77
  77 79
unitary similarity of, 78
upper triangular, 16
  16 22 23 31
determinant of, 50
Vandermonde, 79
zero, see Zero matrix
Modular equation, 12
for codimension, 34
Monomial
multivariate, 69
degree of, 69
monic, 69
monic part of, 69
Multilinearity, 56
of the determinant, 54
Norm
of column vectors, 13, 14, 52, 64, 76, 77
Null space
of a matrix, 34, 38, 38
Nullity
of a matrix, 38
Operator norm
of a matrix, 80, 80, 81
submultiplicativity of, 80
Orthogonal group, 64
Orthogonal matrix
eigenvalues of, 64
Orthogonal system, 13, 76
Orthogonality
of column vectors, 13, 14, 20, 58, 76
with respect to a bilinear form, 73
Orthogonally similar matrices, 64, 66, 73
Orthonormal basis
of a subspace of $\mathbb{R}^n$, 13
of $\mathbb{C}^n$, 77
of $\mathbb{R}^n$, 64, 65
Orthonormal system, 13, 64, 77
$P_n(F)$, 66
Parallelepiped, 52, 53
Parallelogram, 52, 53
Permutation, 23, 23, 25, 48, 48, 49
composition of, 24, 24, 48, 49
decomposition, 49
disjoint, 49, 49
even, 48, 48
fixed point of, 54, 54
identity, 24
inverse of, 24, 24
inversion of, 48, 48, 49
odd, 48
parity, 48, 49
product of, 48, 49
sign, 48, 49
Permutation matrix, 23
Polynomial, 22, 59, 60, 62
degree of, 59, 60
divisibility of, 60
leading coefficient of, 59, 60
leading term of, 59, 60
monic, 59
multivariate, 69, 69, 70
degree of, 70
homogeneous, 70, 70
standard form, 70
of a matrix, 26
of matrices, 24, 24, 24, 62, 62, 63, 119, 120
root of, 60, 60
multiplicity, 60
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>zero, 59 60</td>
<td></td>
</tr>
<tr>
<td>Quadratic form, 71 73</td>
<td></td>
</tr>
<tr>
<td>definiteness of, 71 71 73</td>
<td></td>
</tr>
<tr>
<td>indefinite, 71</td>
<td></td>
</tr>
<tr>
<td>negative definite, 71</td>
<td></td>
</tr>
<tr>
<td>negative semidefinite, 71</td>
<td></td>
</tr>
<tr>
<td>positive definite, 71</td>
<td></td>
</tr>
<tr>
<td>positive semidefinite, 71</td>
<td></td>
</tr>
<tr>
<td>( \mathbb{R}^n )</td>
<td></td>
</tr>
<tr>
<td>orthonormal basis of, 64 65</td>
<td></td>
</tr>
<tr>
<td>subspace of, 7</td>
<td></td>
</tr>
<tr>
<td>Radon’s Lemma, 45</td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td></td>
</tr>
<tr>
<td>of a matrix, 32 32 36 76</td>
<td></td>
</tr>
<tr>
<td>full, 32</td>
<td></td>
</tr>
<tr>
<td>of column vectors, 10 10 11</td>
<td></td>
</tr>
<tr>
<td>Rank-Nullity Theorem</td>
<td></td>
</tr>
<tr>
<td>for matrices, 38</td>
<td></td>
</tr>
<tr>
<td>Rayleigh quotient, 67 71</td>
<td></td>
</tr>
<tr>
<td>Rayleigh’s Principle, 67</td>
<td></td>
</tr>
<tr>
<td>Reflection matrix, 81</td>
<td></td>
</tr>
<tr>
<td>Rook arrangement, 23 23</td>
<td></td>
</tr>
<tr>
<td>Rotation matrix, 18 19</td>
<td></td>
</tr>
<tr>
<td>Row space, 28</td>
<td></td>
</tr>
<tr>
<td>Row vector, 15</td>
<td></td>
</tr>
<tr>
<td>Row-rank, 28</td>
<td></td>
</tr>
<tr>
<td>full, 28 33</td>
<td></td>
</tr>
<tr>
<td>( S_{n} ), see Symmetric group</td>
<td></td>
</tr>
<tr>
<td>Scalar, 3</td>
<td></td>
</tr>
<tr>
<td>Scaling, 29</td>
<td></td>
</tr>
<tr>
<td>Scaling matrix, 30</td>
<td></td>
</tr>
<tr>
<td>Schur’s Theorem, 78</td>
<td></td>
</tr>
<tr>
<td>real version, 79</td>
<td></td>
</tr>
<tr>
<td>Second Miracle of Linear Algebra, 31 31</td>
<td></td>
</tr>
<tr>
<td>Set</td>
<td></td>
</tr>
<tr>
<td>affine closed, 43</td>
<td></td>
</tr>
<tr>
<td>affine-closed, 41 41</td>
<td></td>
</tr>
<tr>
<td>intersection of, 41</td>
<td></td>
</tr>
<tr>
<td>closed, 45</td>
<td></td>
</tr>
<tr>
<td>convex, 44 44 46</td>
<td></td>
</tr>
<tr>
<td>dimension of, 45</td>
<td></td>
</tr>
<tr>
<td>intersection of, 44</td>
<td></td>
</tr>
<tr>
<td>corank of, 34</td>
<td></td>
</tr>
<tr>
<td>open, 43 132</td>
<td></td>
</tr>
<tr>
<td>sum of, 8 8 12</td>
<td></td>
</tr>
<tr>
<td>translate of, 39 39 42</td>
<td></td>
</tr>
<tr>
<td>Shearing, 29 29</td>
<td></td>
</tr>
<tr>
<td>Shearing matrix, 30</td>
<td></td>
</tr>
<tr>
<td>Similar matrices, 58 58 59 62 72 73</td>
<td></td>
</tr>
<tr>
<td>Skew cofactor expansion, 54</td>
<td></td>
</tr>
<tr>
<td>Span</td>
<td></td>
</tr>
<tr>
<td>of column vectors, 7 7 8 10 39</td>
<td></td>
</tr>
<tr>
<td>transitivity of, 8</td>
<td></td>
</tr>
<tr>
<td>Spectral Theorem</td>
<td></td>
</tr>
<tr>
<td>for normal matrices, 79</td>
<td></td>
</tr>
<tr>
<td>for real symmetric matrices, 66 66 67</td>
<td></td>
</tr>
<tr>
<td>applications of, 66 66 67</td>
<td></td>
</tr>
<tr>
<td>Standard basis of ( \mathbb{F}^k ), 10 19 20</td>
<td></td>
</tr>
<tr>
<td>Standard unit vector, 10 19</td>
<td></td>
</tr>
<tr>
<td>Steinitz exchange lemma, 11</td>
<td></td>
</tr>
<tr>
<td>Straight-line segment, 45 45</td>
<td></td>
</tr>
<tr>
<td>Sublist, 9 9</td>
<td></td>
</tr>
<tr>
<td>Submatrix, 25 32 32</td>
<td></td>
</tr>
<tr>
<td>Subset</td>
<td></td>
</tr>
<tr>
<td>dense, 78 78</td>
<td></td>
</tr>
<tr>
<td>Subspace, see Affine subspace</td>
<td></td>
</tr>
<tr>
<td>affine, see Affine subspace</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Page(s)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>of ( \mathbb{F}^k ), ( k )</td>
<td>6, 6, 8, 44</td>
</tr>
<tr>
<td>basis of ( \mathbb{F}^k )</td>
<td>10, 10, 12</td>
</tr>
<tr>
<td>codimension of ( \mathbb{F}^k )</td>
<td>34, 34</td>
</tr>
<tr>
<td>dimension of ( \mathbb{F}^k )</td>
<td>10, 10, 12</td>
</tr>
<tr>
<td>disjoint</td>
<td>12</td>
</tr>
<tr>
<td>intersection of</td>
<td>7</td>
</tr>
<tr>
<td>totally isotropic</td>
<td>74, 74</td>
</tr>
<tr>
<td>union of</td>
<td>7</td>
</tr>
<tr>
<td>zero-weight</td>
<td>7</td>
</tr>
<tr>
<td>Swapping</td>
<td>29, 29</td>
</tr>
<tr>
<td>Swapping matrix</td>
<td>30</td>
</tr>
<tr>
<td>Symmetric group</td>
<td>48, 49</td>
</tr>
<tr>
<td>System of linear equations, ( \text{SL} )</td>
<td>6, 37, 40, 47</td>
</tr>
<tr>
<td>augmented</td>
<td>39</td>
</tr>
<tr>
<td>homogeneous</td>
<td>37, 42, 52</td>
</tr>
<tr>
<td>solution space of</td>
<td>38, 38, 39</td>
</tr>
<tr>
<td>trivial solution to</td>
<td>37</td>
</tr>
<tr>
<td>set of solutions to</td>
<td>39, 42</td>
</tr>
<tr>
<td>solvable</td>
<td>37, 39, 40, 56</td>
</tr>
<tr>
<td>Third Miracle of Linear Algebra</td>
<td>64, 77</td>
</tr>
<tr>
<td>Trace of a matrix</td>
<td>27, 21, 61</td>
</tr>
<tr>
<td>Transposition</td>
<td>28, 48, 49</td>
</tr>
<tr>
<td>neighbor</td>
<td>49</td>
</tr>
<tr>
<td>Triangle inequality</td>
<td>80, 81</td>
</tr>
<tr>
<td>Unitarily similar matrices</td>
<td>78, 78, 79</td>
</tr>
<tr>
<td>Unitary group</td>
<td>77</td>
</tr>
<tr>
<td>Unitary matrix eigenvalues of</td>
<td>77</td>
</tr>
<tr>
<td>Vandermonde determinant</td>
<td>54</td>
</tr>
<tr>
<td>Vandermonde matrix</td>
<td>25, 26, 54</td>
</tr>
<tr>
<td>Vector isotropic</td>
<td>13, 13</td>
</tr>
<tr>
<td>Zero matrix</td>
<td>16, 21, 32</td>
</tr>
<tr>
<td>Zero polynomial</td>
<td>59, 60</td>
</tr>
<tr>
<td>Zero vector, see 0</td>
<td></td>
</tr>
</tbody>
</table>