Federal University of Minas Gerais
Institute of Exact Sciences
Computer Science Department

Algebras, groups and graphs

Henrique Soares Assumpção e Silva

Advisor:
Gabriel Coutinho

Undergraduate Thesis

January 2025

Introduction

This work consists of the Undergraduate Thesis[1] required for the Bachelor's degree in Computer Science at the Federal University of Minas Gerais (UFMG). Our main goal will be to understand how to apply tools from areas of pure mathematics, such as non-commutative algebra and ring theory, to problems in graph theory and combinatorial optimization, which are of great interest to computer science as a whole.

We begin with a *graph* X, which is a set V of vertices and a set $E \subseteq V \times V$ of edges that connect the vertices. Graphs can be found in various research areas: they are used to describe molecules and other chemical structures, to model physical systems of particle interactions, to understand relationships between social network users, and many other things. These structures are also of extreme importance to computer science: 10 of the 21 original NP-complete problems described in [Kar72] are explicitly about graphs, and this diversity of interesting and difficult problems is perhaps one of the main drivers for the development of graph theory throughout the 20th century, which is now an extremely rich and deep area.

There are various ways to create a matrix that captures the combinatorial structure of X, and perhaps the most natural way is through the *adjacency matrix*, denoted by $A(X)$. This is a matrix with rows and columns indexed by the vertices of X, such that the entry uv of the matrix is equal to 1 if uv is an edge of X, and equal to 0 otherwise. From this, we may ask: what kind of combinatorial information about the graph can be extracted from the algebraic properties of its adjacency matrix? Questions like this are characteristic of the areas now known as spectral theory and algebraic graph theory, and our approach will use various techniques from these areas. In particular, we will be interested in certain families of regular graphs whose adjacency matrices are naturally associated with *algebras*. An algebra is nothing more than a vector space equipped with a multiplication operation between its elements, and in general, we will be interested in understanding what kind of combinatorial information we can extract from algebras associated with certain graphs.

In the first chapter, we discuss all the essential prerequisites about groups, rings, and modules, and we prove all the theorems that will be used throughout the work (a reader who is already familiar with ring theory and group theory may skip this chapter). The second chapter discusses semisimple rings and modules, culminating in a proof of the Wedderburn and Artin Theorems. We also prove basic results about the Jacobson radical, and finally, we give a brief discussion of the representation theory of finite groups. The third chapter focuses on complex matrix algebras and presents several constructive and non-constructive proofs about the semisimplicity of certain algebras. In the fourth and final chapter, we discuss the notions of association schemes and coherent configurations, presenting various important examples related to graph theory and finite groups. We also discuss in detail the basics of distance-regular graphs, which are particularly important for algebraic combinatorics, and some applications of association schemes to error-correcting codes theory. As main references, we use chapters $4, 5, 7$ from [Coh12], $1, 2, 4$ from [CR66], 4 from [Far12], $1, 2$ from [FD12], $1, 2, 3$ from [Lam13], $13, 17, 18$ from [Lan05], 4 from [Pas04], $1, 2$ from [VS21], the lecture notes of [Men23] from the UFMG Non-commutative Algebra course, $1, 2$ from [Bai04], $2, 3, 4$ from [BCN11], $15 - 17, 20 - 22$ from [Big93], 2 from [CP23], $10 - 13$ from [God93], $1, 2, 16, 19, 20$ from [God10], $2, 8, 10$ from [GR13]. Additionally, we also discuss results found in other articles and books, which are duly cited when necessary.

We tried to make this material as self-contained as possible, but we assume that the reader is already familiar with the topics of abstract algebra, e.g. [Gal21], linear algebra, e.g. [Axl14], and graph theory, e.g. [Die05].

[1] This is a translated and updated version of the original manuscript published in Portuguese.

Sumário

1 Basic Structures

Our main objective now is to introduce the basic algebraic structures necessary for the other results presented throughout the work. We will begin with groups, and then proceed to rings, modules, and algebras.

1.1 Groups

1.1.1 Basic Concepts

Definition 1.1.1 (Group)**.** Given a set G equipped with a binary operation $\cdot : G \times G \mapsto G$, we say that $(G, \cdot)$ is a *group* if for all $a, b, c \in G$ the following conditions hold:

(1) $a \cdot b \cdot c = a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

(2) There exists an identity element e in G such that $a \cdot e = e \cdot a = a$;

(3) There exists an element a^{-1} in G such that $a \cdot a^{-1} = e = a^{-1} \cdot a$.

Both the identity element and the inverse are unique, and if the operation is commutative, we say the group is *abelian*.

Item (1) tells us that the group's operation is associative, (2) guarantees the existence of an identity element, and (3) shows that every element of a group is invertible. There are several important examples of groups: the *symmetric group* $\mathrm{Sym}(X)$ on a set X consisting of all bijections from X to X with the operation of function composition – and such a set is denoted by $\mathrm{Sym}(n)$ or S_n when X is finite and has n elements; the *general linear group* $\mathrm{GL}(n, \mathbb{C})$ consisting of the invertible $n \times n$ matrices with entries in the complex numbers and the operation of matrix multiplication; the *dihedral group* D_n of symmetries of a regular polygon with n vertices; and many others.

Given a subset $H \subseteq G$, we say that H is a *subgroup* – denoted by $H \leq G$ – if H is also a group with respect to the same operation as G. In this case, we can observe that H is a subgroup of G if and only if H contains xy^{-1} for any $x, y \in H$. There are two common notations for groups: the *multiplicative* notation, where we represent the group operation as multiplication $\cdot$ and its identity element as 1, and the *additive* notation – usually used in abelian groups – where we use $+$ to represent the operation and 0 as the identity element.

If $g, h \in G$, we denote the *conjugation* of g by h as $g^h = hgh^{-1}$, and we say that a subgroup $N \leq G$ is normal if $hNh^{-1} = \{hgh^{-1} | g \in N\} = N$ for any $h \in G$, i.e., if N is invariant under conjugation by elements of G. Given groups G and H, we can study maps between them that preserve their group structure, i.e., functions of the form

$$\varphi : G \mapsto H,$$
$$\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y).$$

These functions are called *group homomorphisms*, and it is immediate to note that $\varphi(e_G) = e_H$, and that $\varphi(x^{-1}) = \varphi(x)^{-1}$. If a homomorphism is injective, it is called a *monomorphism*, if it is surjective, it is called an *epimorphism*, and if it is bijective, it is called an *isomorphism*, and we write $G \cong H$ when G and H are isomorphic groups. A homomorphism from a group to itself is called an *endomorphism*, and when this map is bijective, it is called an *automorphism*, and in this case, we denote the set of automorphisms of a group by $\mathrm{Aut}(G)$, which is also a group under function composition. Given a homomorphism φ between groups G and H, we define its *kernel* and *image* as the sets

$$\ker(\varphi) := \{x \in G | \varphi(x) = e_H\} \subseteq G,$$
$$\mathrm{Im}(\varphi) := \{\varphi(x) | x \in G\} \subseteq H,$$

respectively, and it is immediate to note that $\ker(\varphi)$ is a normal subgroup of G, and that $\mathrm{Im}(\varphi)$ is a subgroup of H. From these definitions, we can observe that a homomorphism is injective if and only if its kernel contains only the identity element, and in this case, we say that the kernel is *trivial*. The terminology we presented for functions between groups will also be used for functions between the other algebraic objects discussed in the following sections, with the necessary adjustments that will be clear from the context.

1.1.2 Actions

We say that G acts on a set X if there exists a homomorphism φ between G and the symmetric group $\mathrm{Sym}(X)$ of X, that is, it is possible to map elements of G to permutations of X. In this case, if $\sigma \in G$ and $x \in X$, we write $\sigma x := \varphi(\sigma)(x)$. A group G always acts naturally on itself via the so-called *regular action* (on the left), defined by

$$\varphi : G \mapsto \mathrm{Sym}(G),$$
$$\varphi(g) = g_L,$$

where $g_L(h) = gh$, that is, we map each element g of the group to the function that multiplies elements of the group by g on the left. We can make an analogous construction for right multiplication via $g \mapsto g_R$ where $g_R(h) = hg^{-1}$, and take the inverse of g to ensure that the map is a homomorphism. Moreover, if $g_L = \mathrm{id}$, then $g = g_L(1) = 1$, so φ is an injective map. In general, a *permutation representation* of G is a homomorphism from G to some symmetric group, and we say that this representation is *faithful* if it is injective, so the previous construction shows that every group admits a faithful permutation representation. These observations lead to the following classical result:

Theorem 1.1.2 (Cayley's Theorem)**.** Every group G has a faithful permutation representation.

■

Whenever G acts on a set X, we can consider sets of the form $Gx = \{\sigma x | \sigma \in G\}$ for any $x \in X$, and notice that the distinct sets of this form partition X. These sets form the so-called *orbits* of the action, and if X_i is an orbit, then $X_i = Gx$ for any $x \in X_i$, and in particular these orbits are precisely the minimal invariant subsets under the action of G, that is, $GX_i = X_i$. We can also consider the set $G_x = \{g \in G | gx = x\}$ of elements of the group that fix x, called the *stabilizer* of x, and we will soon see that the orbits and stabilizers of a group are closely connected. The group G always acts on itself by conjugation, that is, by the map $g \mapsto \varphi(g)$ where $\varphi(g)h = h^g$, and the orbits of this action are called the *conjugacy classes* of the group, while the stabilizer of this action is the so-called centralizer $C_G(g)$ of an element x, which contains all the elements of the group that commute with g. If H is a subgroup of G, then H also acts on G by left and right multiplication, and the orbits of these actions are sets of the form $Hg = \{hg | h \in H\}, gH = \{gh | h \in H\}$, called left and right cosets, respectively. It is worth noting that if $Y \subseteq X$ is a subset invariant under the action of G, then Y is a disjoint union of orbits, and if G acts transitively on Y, then Y is an orbit of the action, that is, a subset is an orbit if and only if it is G-invariant and G acts transitively on it.

If H is a subgroup of G, then the cosets $G/H := \{g_1 H, \ldots, g_k H\}$ of H partition G, but note that $|g_i H| = |g_j H|$, that is, $|G| = k|H|$, where k is the number of cosets of H, and from this we obtain the following famous result:

Theorem 1.1.3 (Lagrange's Theorem)**.** Let G be a finite group and H be a subgroup of G, and let $[G : H]$ be the number of cosets of H in G – called the *index* of H. Then

$$|G| = [G : H] \cdot |H|,$$

and in particular, $|H|$ divides $|G|$.

■

The previous theorem is one of the fundamental basic results in group theory, and now we will see how to use it to obtain another important theorem about orbits and stabilizers.

Theorem 1.1.4 (Orbit-Stabilizer Theorem)**.** If G is a group that acts on a set X, then for each $x \in X$, there is a bijection between the elements of the orbit Gx and the cosets of the stabilizer G_x in X. In particular, if G is finite, then $|Gx| = [G : G_x]$ and

$$|G| = |G_x| \cdot |Gx|.$$

Proof. To prove the theorem, we first fix $x \in X$ and consider its orbit Gx, and then we define the map φ_x that takes elements of Gx to the set of cosets G/G_x,

$$\varphi_x : Gx \mapsto G/G_x$$
$$\varphi_x(y) = gG_x,$$

where $g \in G$ is such that $gx = y$. First, we show that the function is well-defined, that is, if $g_1, g_2 \in G$ are such that $g_1x = g_2x = y$, we want to show that $g_1G_x = g_2G_x$. Indeed, we have that

$$g_2^{-1}g_1x = g_2^{-1}y = x,$$

that is, $g_2^{-1}g_1 \in G_x$, and therefore $g_1G_x = g_2G_x$. Note also that if $g_1G_x = g_2G_x$, then $g_2^{-1}g_1 \in G_x$, so $g_1x = g_2x$, which shows the injectivity of the map. Finally, if gG_x is a coset, we just consider the element $gx \in Gx$ from the orbit and notice that $\varphi_x(gx) = gG_x$, and hence φ_x is a bijection. In the case of a finite group, this bijection implies that $|Gx|$ is finite and equal to $[G : G_x]$, so by Lagrange's Theorem we have that

$$|G| = |G_x| \cdot [G : G_x] = |G_x| \cdot |Gx|,$$

as we wanted. ■

We say that G acts transitively on X if for any $x, y \in X$, there exists $g \in G$ such that $gx = y$, that is, for any $x \in X$, the orbit $Gx = X$, and thus the previous theorem tells us that in the case of transitive groups:

$$|G| = |G_x| \cdot |X|,$$

that is, $|X|$ divides the order of the group.

1.1.3 Products and Sums

A construction that will be important for the upcoming sections is the notion of products between groups and subgroups. If H, K are subgroups of G, we define the product HK as

$$HK = \{hk | h \in H, k \in K\},$$

and it is immediate to check that HK is a subgroup if and only if $HK = KH$, and it is also worth mentioning that if K is normal, then $HK = KH$ and thus the product is a subgroup. We can also form a new group from distinct groups:

Definition 1.1.5 (External Direct Product of Groups)**.** Let G, H be groups, then we define their *external direct product* as the set

$$G \times H := \{(x, y) | x \in G, y \in H\},$$

with operation given by

$$(x_1, y_1)(x_2, y_2) := (x_1x_2, y_1y_2).$$

This set is a group with identity element given by (e_G, e_H).

If $G_1 \times G_2$ is an external direct product of groups, we can consider the normal subgroups $G_1' = \{(g_1, e_2) | g_1 \in G_1\}, G_2' = \{(e_1, g_2) | g_2 \in G_2\}$, and notice that $G_1 \times G_2 = G_1'G_2'$, and that $G_1' \cap G_2' = \{(e_1, e_2)\}$. On the other hand, if N, M are normal subgroups of G such that $NM = G$ and $N \cap M = \{e\}$, then $G \cong N \times M$, and in practical terms, this means that every element of the group can be uniquely expressed as a product nm, with $n \in N, m \in M$. That is, there is an equivalence between the internal product of groups and the external direct product, and these two different ways of viewing the product will be useful in future chapters.

We can also generalize the notion of direct product for an arbitrary number of groups, that is, if $\{G_i\}_{i \in \mathcal{I}}$ is an arbitrary family of groups indexed by some set $\mathcal{I}$, we define their external direct product as

$$\prod_{i \in \mathcal{I}} G_i := \{(x_i)_{i \in \mathcal{I}} | x_i \in G_i\},$$

with operation given by

$$(x_i)_{i\in\mathcal{I}}(y_i)_{i\in\mathcal{I}} := (x_i y_i)_{i\in\mathcal{I}},$$

for any sequences $(x_i)_{i\in\mathcal{I}}, (y_i)_{i\in\mathcal{I}}$ of $\prod_{i\in\mathcal{I}} G_i$. The direct product is a group where the identity element is just the sequence with each identity element of the respective G_i. It is also naturally accompanied by a family of group homomorphisms: for each index i of $\mathcal{I}$, we can define the *projection* epimorphism that maps any sequence in the product to its i-th element, and the *inclusion* monomorphism, which maps an element x of G_i to the sequence formed by the identity elements in positions different from i and by x in the i-th position. In the case of abelian groups, we also define the notion of *external direct sum*. If $\{G_i\}_{i\in\mathcal{I}}$ is a family of abelian groups, we denote by

$$\bigoplus_{i\in\mathcal{I}} G_i \subseteq \prod_{i\in\mathcal{I}} G_i$$

the subset of the direct product formed by all *almost-zero* sequences, that is, sequences $(a_i)_{i\in\mathcal{I}}$ that have only a finite number of elements different from the identity element of the respective group. In this case, this set forms a group with the same operations and identity element as the external direct product, and it is worth noting that if $\mathcal{I}$ is finite, then the direct product and the direct sum are equal.

Finally, we will discuss a last group product that will be used in some applications, known as the *semidirect product*. If $G = NH$, where N is a normal subgroup and H is any subgroup, and $N \cap H = \{e\}$, then we say that G is the internal semidirect product of N and H, and we write

$$G = N \rtimes H = H \ltimes N.$$

There are several examples of semidirect products, such as the dihedral group D_n which can be written as $D_n = C_n \rtimes C_2$, where C_n is the subgroup formed by rotations, and C_2 is the subgroup formed by reflections, or the group UT_n of upper triangular matrices with non-zero determinant, which can be written as $\mathbb{UT}_n = \mathbb{U}_n \rtimes \mathbb{D}_n$, where $\mathbb{U}_n$ is the subgroup of upper triangular matrices with 1's on the diagonals, and $\mathbb{D}_n$ is the subgroup of diagonal matrices with non-zero entries. We can notice that if $G = N \rtimes H$, given $n_1h_1, n_2h_2 \in G$, we have

$$n_1h_1n_2h_2 = n_1n_2^{h_1}h_1h_2,$$

so if we define the homomorphism φ from H to $\mathrm{Aut}(N)$ such that $\varphi(h)(n) = n^h$, we get

$$n_1h_1n_2h_2 = n_1\varphi(h_1)(n_2)h_1h_2,$$

where $h_1h_2 \in H$, and $n_1\varphi(h_1)(n_2) \in N$. This observation motivates the definition of external semidirect products. If H, N are groups such that there exists a homomorphism φ from H to $\mathrm{Aut}(N)$, we define the *external semidirect product* as

$$N \rtimes H := \{(x, y) | x \in N, y \in H\},$$

with the operation given by

$$(n_1, h_1)(n_2, h_2) = (n_1\varphi(h_1)(n_2), h_1h_2),$$

and the identity element given by (e_N, e_H). Again, we will have an equivalence between internal and external semidirect products, but these two interpretations have different utilities that we will use in the future.

1.2 Rings and Fields

1.2.1 Basic Concepts

The main structure of interest for this work is the so-called rings, which are sets where one can add and multiply elements in an associative manner. Rings are also called number systems, and historically the motivation behind their study comes from the use of number systems alternative to the integers – such as Gaussian integers – to prove results in number theory. Formally, we have the following definition:

Definition 1.2.1 (Ring)**.** Let $(R, +)$ be an abelian group. We say that R is a *ring* if it is equipped with an operation $\cdot : R \times R \mapsto R$ such that for any elements a, b, c of R:

(1) $a \cdot b \cdot c = a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

(2) $a \cdot (b + c) = a \cdot b + a \cdot c$;

(3) $(a + b) \cdot c = a \cdot c + b \cdot c$;

(4) There exists an element 1 in R such that $a \cdot 1 = 1 \cdot a = a$.

As in the case of groups, it is immediate to verify that the multiplicative identity of the ring is unique. We will denote $a \cdot b$ as ab when convenient, and if the operation $\cdot$ is commutative, R is said to be *commutative*.

Item (1) tells us that the multiplication operation of the ring is associative, (2) and (3) tell us that the multiplication operation is distributive with respect to addition, and (4) guarantees the existence of a multiplicative identity element[2]. The set $\mathbb{Z}$ of integers is a classical example of a commutative ring. The set $M_n(R)$ of $n \times n$ matrices with entries in a given ring R is also an important example of a ring, and note that even if R is commutative, the ring $M_n(R)$ is not commutative if $n \geq 2$. This ring is also called the *full matrix algebra* of $n \times n$ matrices with entries in R, for reasons that will become clear in the following sections.

Given non-zero elements a, b of R such that $ab = 1$, we say that a is *right invertible*, and similarly that b is *left invertible*. If an element is invertible both on the left and on the right, then the inverses are equal, and we say it is *invertible*. The set of invertible elements of a ring R is denoted by R^* or $U(R)$, and is usually called the *unit group of the ring*, since it is indeed a group under multiplication with the identity element given by 1. If every non-zero element of a ring is invertible, that is, if $R^* = R \setminus \{0\}$, we say that R is a *division ring*, and if it is also commutative, it is called a *field*.

Example 1.2.2. Let $\mathbb{Z}_n$ denote the set of equivalence classes of integers modulo n. It is known that an element $x \in \mathbb{Z}_n$ has an inverse if and only if x is coprime to n, i.e., $\mathbb{Z}_n^* := \{x \in \mathbb{Z}_n | \gcd(x, n) = 1\}$ is the unit group of $\mathbb{Z}_n$, also called the multiplicative group of the ring. It is worth noting that the cardinality of $\mathbb{Z}_n^*$ is given by $\varphi(n)$, where φ is Euler's totient function. If p is prime, then every non-zero element will be invertible, and therefore $\mathbb{Z}_p$ is a field.

Given any ring R and an abelian subgroup S of R, we say that S is a *subring* if S is also a ring with the same operations as R and contains the identity element 1 of R. An abelian subgroup L is a *left ideal* of R if it is closed under multiplication from the left by elements of R, that is, the product ra belongs to L for any elements r of R and a of L. An analogous definition applies for right ideals, and if an ideal is both a left and a right ideal, it is called *bilateral* (or simply an ideal), and naturally, if R is commutative, every ideal is bilateral. It follows immediately that if L is a left ideal that contains the identity 1, then $R = L$, i.e., any proper non-trivial ideal of R does not contain the identity and therefore is not a subring. A ring R is said to be *simple* if its only bilateral ideals are R and $\{0\}$. A proper left ideal L is said to be *maximal* if, for any other left ideal J such that $L \subseteq J$, it holds that $J = L$ or $J = R$, and a left ideal L is said to be *minimal* if the only proper left ideal of L is the trivial ideal $\{0\}$. Analogous definitions apply for right ideals and bilateral ideals.

A homomorphism φ between two rings R, S is a group homomorphism between $(R, +)$ and $(S, +)$ that is also compatible with the multiplication operations of the rings, i.e.,

$$\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b),$$
$$\varphi(1_R) = 1_S,$$

for any elements a, b of R. We will denote the set of homomorphisms between rings R and S by $\mathrm{Hom}(R, S)$, and if $R = S$, we will denote this same set by $\mathrm{End}(R)$, which in this case is also a ring with the composition of functions as the operation and addition, with the identity given by the identity function.

An extremely common construction in the study of abstract algebra is the so-called quotients. In the case of rings, given an ideal I of a ring R, we create a new ring denoted by

$$R/I := \{a + I | a \in R\}$$

[2]For this work, a ring is always assumed to be associative and with identity. It is possible to discuss non-unital or non-associative rings, but these will not be of interest for our work.

called the *quotient ring*. The elements $a + I := \{a + b | b \in I\}$ are called *cosets*, and may also be denoted by $\overline{a}$. The addition and multiplication operations are defined as follows:

$$(a + I) + (b + I) := (a + b) + I;$$
$$(a + I)(b + I) := ab + I.$$

The additive identity of R/I is given by $0 + I = I$ – which will also be denoted by 0 when the context is clear –, and the multiplicative identity is given by $1 + I$. We leave it to the reader to verify that R/I is a ring if, and only if, I is a bilateral ideal of R, i.e., we cannot take quotients over just any ideal of a given ring.

We can also consider the set of all elements that commute with all others in R, usually called the *center of the ring*, and denoted by

$$Z(R) := \{a \in R | \forall b \in R : ab = ba\} \subseteq R,$$

which naturally forms a subring of R. In this case, we can also define the *commutator* between two elements of R as

$$[a, b] := ab - ba,$$

and thus we can describe the center of the ring as

$$Z(R) = \{a \in R | \forall b \in R : [a, b] = 0\}.$$

If $S \subseteq R$ is any subset of a ring R, we define the *centralizer* – also called the commutant – of S as the set of elements of R that commute with all elements of S, i.e.,

$$C_R(S) := \{a \in R | \forall b \in S : [a, b] = 0\} \subseteq R,$$

and note that $C_R(S)$ is a subring of R.

1.2.2 Isomorphism and Correspondence Theorems

There are two extremely important results related to quotient rings, and analogous versions of these results will also hold for modules, which will be discussed in the next section.

Theorem 1.2.3 (First Isomorphism Theorem for Rings)**.** Let R, S be rings, and let $\varphi \in \mathrm{Hom}(R, S)$ be a homomorphism between them. Then $\ker(\varphi)$ is a bilateral ideal of R, $\mathrm{Im}(\varphi)$ is a subring of S, and

$$R/\ker(\varphi) \cong \mathrm{Im}(\varphi).$$

Proof. We leave it to the reader to prove that $\ker(\varphi)$ is a bilateral ideal of R and that $\mathrm{Im}(\varphi)$ is a subring of S. Define the following function:

$$\overline{\varphi} : R/\ker(\varphi) \mapsto \mathrm{Im}(\varphi),$$
$$\overline{\varphi}(r + \ker(\varphi)) = \varphi(r),$$

and note that $\overline{\varphi}$ is a ring homomorphism.

Let r be an element of R and assume that $\overline{\varphi}(r + \ker(\varphi)) = 0$, so by definition $\varphi(r) = 0$, meaning r is an element of the kernel $\ker(\varphi)$ of φ, which is the zero element of $R/\ker(\varphi)$. Therefore, $\ker(\overline{\varphi}) = \{0\}$, implying that $\overline{\varphi}$ is injective. Given any element $\varphi(r)$ in the image $\mathrm{Im}(\varphi)$, where r belongs to R, note that $r + \ker(\varphi)$ belongs to $R/\ker(\varphi)$, and that $\overline{\varphi}(r + \ker(\varphi)) = \varphi(r)$, so $\overline{\varphi}$ is surjective. From these observations, it follows that $\overline{\varphi}$ is an isomorphism of rings, allowing us to conclude that $R/\ker(\varphi)$ is indeed isomorphic to $\mathrm{Im}(\varphi)$. ■

The previous theorem essentially tells us that for any homomorphism φ between rings R, S, there always exists a homomorphism $\overline{\varphi}$ such that the diagram

$$\begin{array}{ccc} R & \xrightarrow{\quad\varphi\quad} & S \\ \downarrow{\scriptstyle \pi} & & \uparrow{\scriptstyle \iota} \\ R/\ker(\varphi) & \xrightarrow{\quad\overline{\varphi}\quad} & \mathrm{Im}(\varphi) \end{array}$$

commutes, that is, such that $\varphi = \iota \circ \overline{\varphi} \circ \pi$, where $\pi(a) = a + \ker(\varphi), \iota(b) = b$ denote the canonical projection and inclusion maps, respectively. It is also important to note that if φ is injective, the previous theorem implies that R is isomorphic to $\mathrm{Im}(\varphi)$, and in this case, we say that R is *isomorphically immersed* in S, because we can identify R as a subring of S.

Now we show that we can identify the left ideals of R with the left ideals of its quotient ring with respect to some ideal.

Theorem 1.2.4 (Correspondence Theorem for Rings)**.** Let R be a ring, and let I be a bilateral ideal of R. Then there is a bijection between the left ideals of R that contain I and the left ideals of R/I.

Proof. Consider the sets

$$\mathcal{L}_R := \{L \subseteq R \mid L \text{ is a left ideal of } R, I \subseteq L\},$$
$$\mathcal{L}_{R/I} := \{J \subseteq R/I \mid J \text{ is a left ideal of } R/I\},$$

and define the functions

$$f : \mathcal{L}_R \mapsto \mathcal{L}_{R/I} \qquad g : \mathcal{L}_{R/I} \mapsto \mathcal{L}_R,$$
$$f(L) = \{\pi(x) \mid x \in L\} \qquad g(J) = \{x \in R \mid \pi(x) \in J\},$$

where π is the canonical projection mapping x to $x + I$. We will show that f and g are inverse functions, and then conclude the theorem. First, we must show that the functions are well-defined. Consider L a left ideal of R in $\mathcal{L}_R$, and take elements $x + I, y + I$ in the image of $f(L)$ under f, so

$$(x + I) - (y + I) = (x - y) + I = \pi(x - y) \in f(L),$$
$$(r + I)(x + I) = (rx) + I = \pi(rx) \in f(L),$$

and it follows that $f(L)$ belongs to $\mathcal{L}_{R/I}$. Similarly, note that if x is an element of I, then $\pi(x) = I = 0$ in the quotient, and it follows that $g(J)$ belongs to $\mathcal{L}_R$, so the functions are well-defined. Now note that if L is an element of $\mathcal{L}_R$, then

$$g(f(L)) = g(\{\pi(x) \mid x \in L\}) = L,$$

and similarly we have $f(g(J)) = J$ for any J in $\mathcal{L}_{R/I}$, so f and g are inverses of each other, and it follows that f is a bijection. This implies that the left ideals of R that contain I correspond uniquely to the left ideals of R/I. ■

Note that in the previous proof, the function f corresponds exactly to the projection function applied to the ideal L, and the function g corresponds to the pre-image of π with respect to some ideal J of the quotient. It is important to observe that the correspondence can be extended to maximal ideals, that is, each maximal left ideal of R that contains I corresponds to some maximal left ideal of R/I. In particular, if R, S are isomorphic rings, then the previous theorem tells us that the left ideals of R correspond to the left ideals of S. In the case of bilateral ideals, we obtain that every bilateral ideal of R/I is of the form J/I for some bilateral ideal J in R that contains I, and conversely, J/I is a bilateral ideal of R/I for any bilateral ideal J of R that contains I.

1.2.3 Products and Sums

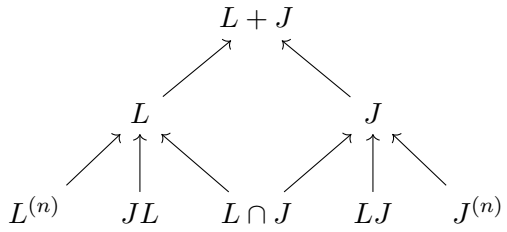
There are some natural operations that can be performed with rings and their ideals. Given left ideals L, J of R, we can define their sum as

$$L + J := \{a + b \mid a \in L, b \in J\} \subseteq R,$$

and note that the sum is also a left ideal. If $\{L_i\}_{i \in \mathcal{I}}$ is a family of left ideals indexed by a set $\mathcal{I}$, we denote by $\sum_{i \in \mathcal{I}} L_i$ the set of finite sums of elements from the respective L_i. We can also define the product of ideals as

$$LJ := \{\sum_i a_i b_i \mid a_i \in L, b_i \in J\} \subseteq R,$$

where all the sums involved are finite, and in this case LJ is a left ideal of R. Therefore, for any natural number n, we can define the left ideal $L^{(n)}$ as the product of L with itself n times, i.e., the set of all possible finite sums of products of n elements from L. The following diagram illustrates the relationships between the different operations we can perform with ideals:



where each directed arrow indicates set inclusion.

It is also possible to extend the notion of direct products of groups to rings:

Definition 1.2.5 (External Direct Product of Rings)**.** Let $\{R_i\}_{i\in\mathcal{I}}$ be an arbitrary family of rings indexed by a set $\mathcal{I}$. Then we define the *external direct product* as

$$\prod_{i\in\mathcal{I}} R_i := \{(a_i)_{i\in\mathcal{I}} \mid a_i \in R_i\},$$

with operations given by

$$(a_i)_{i\in\mathcal{I}} + (b_i)_{i\in\mathcal{I}} := (a_i + b_i)_{i\in\mathcal{I}},$$
$$(a_i)_{i\in\mathcal{I}}(b_i)_{i\in\mathcal{I}} := (a_i b_i)_{i\in\mathcal{I}},$$

for any sequences $(a_i)_{i\in\mathcal{I}}, (b_i)_{i\in\mathcal{I}}$ in $\prod_{i\in\mathcal{I}} R_i$. The additive identity of the direct product is given by $(0_i)_{i\in\mathcal{I}}$, and its multiplicative identity is given by $(1_i)_{i\in\mathcal{I}}$.

It is worth noting that the canonical projections that map a given sequence in the direct product to its i-th element are ring epimorphisms, however, the canonical inclusions defined for groups are not. This follows from the fact that, by definition, a ring homomorphism must map the identity element of the domain to the identity element of the codomain, and this is not the case for inclusions, as they map the identity 1_i of a ring R_i to the sequence with 1_i in the i-th position and 0_j in the others, and this is not the identity element of the direct product. For this reason, in this work, we do not discuss direct sums of rings, even though in some contexts—particularly when dealing with rings that do not have a unit—such objects make sense.

Given a ring $(R, +, \cdot)$, we define its *opposite ring* $(R^{\mathrm{op}}, +, *)$ as the ring with the same addition operation as R, but where for any elements a, b in R, the product is given by

$$a * b := b \cdot a,$$

that is, multiplication in the opposite ring naturally occurs in the reverse order. From this definition, we can demonstrate some basic facts about the opposite ring.

Proposition 1.2.6. If R is a ring and R^{op} is its opposite ring, then:

(1) $(R^{\mathrm{op}})^{\mathrm{op}}$ and R are isomorphic rings;

(2) If R is a division ring, then R^{op} is also a division ring;

(3) If $\{R_i\}_{i\in\mathcal{I}}$ is a family of rings, then $(\prod_{i\in\mathcal{I}} R_i)^{\mathrm{op}}$ and $\prod_{i\in\mathcal{I}} R_i^{\mathrm{op}}$ are isomorphic rings;

(4) Given a natural number n, $M_n(R)^{\mathrm{op}}$ and $M_n(R^{\mathrm{op}})$ are isomorphic rings.

Proof.

(1) It is enough to note that if $*_1$ and $*_2$ are the multiplications in R^{op} and $(R^{\text{op}})^{\text{op}}$, respectively, then

$$a *_2 b = b *_1 a = ab,$$

so the identity function between R and $(R^{\text{op}})^{\text{op}}$ will be a ring isomorphism.

(2) If R is a division ring, we know that any non-zero element a is invertible, so there exists b such that $ab = 1 = ba$, hence $b * a = 1 = a * b$, and therefore a is also invertible in R^{op}.

(3) As in (1), the identity function will be a ring isomorphism. In fact, since the product of two elements in $\prod_{i\in\mathcal{I}} R_i$ is given by the sequence of term-by-term products of each component, the product in the opposite ring will be given by the sequence of term-by-term products in the opposite ring of each component.

(4) Consider the transpose application

$$\varphi : M_n(R)^{\text{op}} \mapsto M_n(R^{\text{op}}),$$
$$\varphi(A) = A^T.$$

The function is clearly an isomorphism of abelian groups and $\varphi(I) = I$, and we can also observe that if A, B are elements of $M_n(R)^{\text{op}}$, then

$$\varphi(A * B) = \varphi(BA) = (BA)^T = A^T B^T = \varphi(A)\varphi(B),$$

which implies that the rings are indeed isomorphic.

■

1.3 Modules and Vector Spaces

1.3.1 Basic Concepts

We have seen that rings generalize the notion of fields, and similarly, we are now interested in studying objects that generalize the notion of a vector space over a field.

Definition 1.3.1 (Module)**.** Let $(M, +)$ be an abelian group and R a ring. We say that M is a left R-module if there exists an operation $\cdot : R \times M \mapsto M$ such that for any elements $m_1, m_2 \in M$, and for any elements $\alpha, \beta \in R$, the following hold:

(1) $\alpha \cdot (m_1 + m_2) = \alpha \cdot m_1 + \alpha \cdot m_2$;

(2) $(\alpha + \beta) \cdot m_1 = \alpha \cdot m_1 + \beta \cdot m_1$;

(3) $\alpha\beta \cdot m_1 = \alpha \cdot (\beta \cdot m_1)$;

(4) $1 \cdot m_1 = m_1$.

In this case, it follows that $0m = 0 = \alpha 0$ and $(-1)m = -m$. A right R-module is defined analogously, and an R-module that is both left and right is called *bilateral*.

Items (1) and (2) tell us that $\cdot$ is compatible with the addition operations of the abelian group M and the ring R, (3) tells us that the operation is associative with respect to the ring multiplication, and (4) simply tells us that the unit of R also acts as the unit in M. The usual Euclidean space $\mathbb{R}^n$ is a bilateral $\mathbb{R}$-module, and it is also a left module with respect to the ring of matrices $M_n(\mathbb{R})$. In general, any module over a commutative ring is bilateral. If G is any abelian group, we can treat it as a $\mathbb{Z}$-module by defining the multiplication of an element of the group by an integer n simply as the sum of that element with itself n times, and if n is negative, we sum the additive inverse of the element. Any ring R is a bilateral module over itself, that is, R is always an R-bilateral module, and in particular, we can always treat a ring R as a left R-module.

Given an abelian subgroup N of an R-module M, we say that it is a *submodule* if N is also an R-module with the same operations as M. An R-module M is called *simple* or *minimal* if its only submodules are M and $\{0\}$. A proper submodule N of M is called *maximal* if, given any other submodule N' such that $N \subseteq N'$, we have $N' = M$ or $N' = N$.

Example 1.3.2 (Simplicity of $M_n(D)$)**.** Let n be a natural number and D a division ring, and consider the ring $M_n(D)$. Note that if I is a bilateral ideal, then we can take a non-zero matrix A from I and multiply it on the left and right by appropriate matrices of the form E_{xy}, that is, with entry xy equal to 1 and the others equal to zero, so as to obtain a matrix with only one non-zero element. Then we can permute the entries of this new matrix to obtain a diagonal matrix with only one non-zero element, and since the ring is a division ring, we can obtain a diagonal matrix with only one non-zero entry equal to 1. Since all operations were just multiplications by matrices in $M_n(D)$, we have shown that I contains all the diagonal matrices of the form E_{xx}, and therefore contains their sum, that is, it contains the identity of $M_n(D)$. This shows that any proper ideal of $M_n(D)$ is trivial, i.e., the ring in question is simple.

From the definitions given, it follows that the R-submodules of R are precisely its left ideals. Therefore, any minimal left ideal of R is a simple R-submodule of R, and conversely, any simple R-submodule of R is a minimal left ideal. Simple modules are objects of great interest in the study of non-commutative algebra, because in many cases we can write a module as a composition of its simple submodules, more precisely, as a direct sum, and throughout the next chapters, we will be interested in determining when such a decomposition is possible.

1.3.2 Isomorphisms

Now we will consider functions φ between R-modules M, N that preserve their structure, that is, group homomorphisms where

$$\varphi(\alpha m) = \alpha\varphi(m),$$

for any elements $\alpha \in R$ and $m \in M$. These functions are called left R-homomorphisms, and the set of all such functions will be denoted by $\mathrm{Hom}_R(M, N)$. Note that this set is always an abelian group with function addition and the additive identity given by the function that is identically zero, but it is an R-module only when R is commutative, and the set $\mathrm{End}_R(M) = \mathrm{Hom}_R(M, M)$ is a ring with addition and function composition as operations.

Given an R-module M and an R-submodule N, we can also consider the quotient module M/N formed by the cosets $m + N$ where $r(m + N) = rm + N$, and in this case, we can observe that the canonical projection mapping an element m to its coset in the quotient will be an epimorphism of modules. If R is a ring and L is a left ideal, we can then consider the quotient R-module R/L, because L is also an R-submodule of R, but it is worth noting that such a module will be a ring if and only if L is bilateral. Since ideals are not subrings, we will compare them through module homomorphisms, that is, two ideals of a ring are isomorphic if there exists a module isomorphism between them. If M, N are isomorphic as left R-modules, we will indicate this by the notation $M \cong_R N$. The following theorems are analogous versions of Theorems 1.2.3 and 1.2.4 for R-modules:

Theorem 1.3.3 (First Isomorphism Theorem for Modules)**.** Let M, N be modules over a ring R, and let $\varphi \in \mathrm{Hom}_R(M, N)$ be a homomorphism between them. Then $\ker(\varphi)$ is a submodule of M, $\mathrm{Im}(\varphi)$ is a submodule of N, and

$$M/\ker(\varphi) \cong_R \mathrm{Im}(\varphi).$$

■

Theorem 1.3.4 (Correspondence Theorem for Modules)**.** Let M be a module over a ring R and N a submodule of M. Then there is a bijection between the left R-submodules of M that contain N and the left R-submodules of M/N. In particular, each R-submodule of M/N is of the form N'/N for some submodule N' of M that contains N, and N'/N is a submodule of M/N for any submodule N' of M. ■

It is important to emphasize that the previous correspondence also holds for maximal submodules of M: if N' is a maximal submodule of M containing N, then N'/N is a maximal submodule of M/N, and vice versa. It is also possible to show that if M, N are isomorphic modules, then such an isomorphism induces a ring isomorphism between $\mathrm{End}_R(M)$ and $\mathrm{End}_R(N)$.

Proposition 1.3.5. If M and N are isomorphic R-modules, then the rings $\mathrm{End}_R(M)$ and $\mathrm{End}_R(N)$ are isomorphic.

Proof. Let φ be an R-isomorphism between M and N, and define:

$$\begin{aligned}\psi : \mathrm{End}_R(M) &\mapsto \mathrm{End}_R(N),\\ \psi(f) &= \varphi \circ f \circ \varphi^{-1}.\end{aligned}$$

Note that $\psi(\mathrm{id}_M) = \mathrm{id}_N$ – where id is the identity function on the respective module –, $\psi(f+g) = \psi(f)+\psi(g)$, and that

$$\begin{aligned}\psi(f \circ g) &= \varphi \circ f \circ g \circ \varphi^{-1}\\ &= (\varphi \circ f \circ \varphi^{-1})(\varphi \circ g \circ \varphi^{-1})\\ &= \psi(f) \circ \psi(g),\end{aligned}$$

so ψ is a ring homomorphism. If $\psi(f) = 0$, this implies that for any element $n \in N$,

$$(\varphi \circ f \circ \varphi^{-1})(n) = 0,$$

but since φ is a bijection, we have for any element $m \in M$

$$(\varphi \circ f)(m) = 0,$$

and again using the fact that φ is a bijection, we obtain that f is identically zero, so ψ is injective. If g is an element of $\mathrm{End}_R(N)$, we can define the function $\tilde{g} = \varphi^{-1} \circ g \circ \varphi$ and note that it is clearly an R-endomorphism of M, and that $\psi(\tilde{g}) = g$. The two previous observations guarantee that ψ is indeed a ring isomorphism, as desired. ■

1.3.3 Products and Sums

Given a subset S of an R-module M, we define the submodule generated by S as the set of finite sums of the form

$$RS := \{\sum_j \alpha_{i_j} m_{i_j} \mid \alpha_{i_j} \in R, m_{i_j} \in S\} \subseteq M.$$

If S generates M, it is called a *generator set*, and if it is finite, we say that M is *finitely generated*. If S contains only one element m, M is called *cyclic*, and is usually denoted by $M = Rm$. If L is a left ideal of R, then the set

$$LS := \{\sum_j \alpha_{i_j} m_{i_j} \mid \alpha_{i_j} \in L, m_{i_j} \in S\} \subseteq M$$

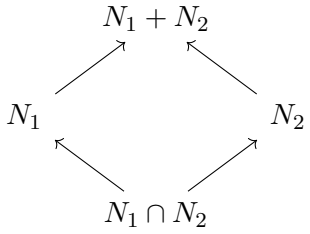
is also an R-submodule of M. If we treat R as a left R-module and if $S \subseteq R$ is any subset, then the left ideal generated by S is precisely the set RS, and if L is a left ideal of R generated by a single element, we say it is a *principal ideal*.

Given an R-module M and a set $S \subseteq M$, we say that S is *linearly independent* if $\sum_{i_j} \alpha_{i_j} m_{i_j} = 0$ implies that $\alpha_{i_j} = 0$ for all i, j, where $\alpha_{i_j} \in R$, $m_{i_j} \in S$; that is, any finite linear combination of elements of S that results in the zero element must have all coefficients zero. A set S is said to be *linearly dependent* if it is not linearly independent, i.e., there exists a finite linear combination that results in the zero element where not all coefficients are zero. A linearly independent generating set is called a *basis* of the R-module, and in this case, we say that the module is *free*, and it is worth noting that every ring R is a free R-module with the basis given by the unit.

If N_1, N_2 are submodules of M, we can define their sum as

$$N_1 + N_2 := \{m + m' \mid m \in N_1, m' \in N_2\} \subseteq M,$$

which is also an R-submodule, and if $\{M_i\}_{i\in\mathcal{I}}$ is a family of submodules of M indexed by a set $\mathcal{I}$, we can consider their sum $\sum_{i\in\mathcal{I}} M_i$ as the set of all finite sums of elements from M_i, which will also be an R-submodule. Again, we can illustrate the relationship between these sets with a diagram:



where the arrows indicate set inclusion.

The concept of direct products of groups can be naturally extended to products of R-modules:

Definition 1.3.6 (Direct product of modules)**.** If $\{M_i\}_{i\in\mathcal{I}}$ is a family of R-modules indexed by a set $\mathcal{I}$, we define the *external direct product*

$$\prod_{i\in\mathcal{I}} M_i := \{(m_i)_{i\in\mathcal{I}} \mid m_i \in M_i\},$$

with operations given by

$$(m_i)_{i\in\mathcal{I}} + (n_i)_{i\in\mathcal{I}} := (m_i + n_i)_{i\in\mathcal{I}},$$
$$\alpha(m_i)_{i\in\mathcal{I}} := (\alpha m_i)_{i\in\mathcal{I}},$$

for any sequences $(m_i)_{i\in\mathcal{I}}, (n_i)_{i\in\mathcal{I}}$ in $\prod_{i\in\mathcal{I}} M_i$ and any $\alpha \in R$. This set is an R-module, and in this case, note that the canonical projections and inclusions will indeed be R-module homomorphisms.

In the case of modules, we can define the *external direct sum* in a similar way to how we did for abelian groups: simply consider the subset of the external direct product formed by the almost-zero sequences with the same operations as the direct product. This set will be an R-module, and is denoted by $\bigoplus_{i\in\mathcal{I}} M_i$. If n is a natural number and M is an R-module, we denote by M^n the external direct sum of M with itself n times.

The following proposition provides an extremely useful criterion for identifying sums of submodules with external direct sums:

Proposition 1.3.7. If M is an R-module and $\{M_i\}_{i\in\mathcal{I}}$ is a family of submodules, the following are equivalent:

(1) $M = \sum_{i\in\mathcal{I}} M_i$ and $M_j \cap (\sum_{i\neq j} M_i) = \{0\}$ for any $j \in \mathcal{I}$;

(2) Every element $m \in M$ can be uniquely written as a finite sum $\sum_j m_{i_j}$, where each m_{i_j} is an element of the corresponding M_{i_j};

(3) $M \cong_R \bigoplus_{i\in\mathcal{I}} M_i$.

Proof.

(1) $\Rightarrow$ (2) Let m be an element of M, and suppose it can be written in two ways as a finite sum of elements from the given ideals. We can assume that

$$m = \sum_j m_{i_j} = \sum_j m'_{i_j},$$

since we can set $m'_{i_j} = 0$ or $m_{i_j} = 0$ if necessary to ensure the index sets are equal. Now, if $\sum_j (m_{i_j} - m'_{i_j}) = 0$, it follows that:

$$(m'_{i_j} - m_{i_j}) = \sum_{l\neq j} (m_{i_l} - m'_{i_l}) \in M_{i_j} \cap (\sum_{l\neq j} M_l),$$

so by assumption, we have $m_{i_j} = m'_{i_j}$, and repeating the same argument for all indices i_j we conclude that the expression for m is indeed unique.

(2) $\Rightarrow$ (3) If each element of M can be uniquely written as a finite sum of elements in the submodules of the family, it is sufficient to consider the map that takes an element $m = \sum_j m_{i_j}$ to the sequence with entries m_{i_j} at the corresponding indices, and zero at the others. It is immediate to verify that this map is indeed an isomorphism of R-modules, and therefore the result follows.

(3) $\Rightarrow$ (1) It is enough to observe that if m is an element of M and is identified with the almost-zero sequence $(m_i)_{i\in\mathcal{I}}$ by the module isomorphism, the element $\sum_{i\in\mathcal{I}} m_i$ which is finite is also identified with this same sequence, hence $M = \sum_{i\in\mathcal{I}} M_i$. Furthermore, if there is an element m_j of M_j such that

$$m_j = \sum_{i\neq j} m_i \in M_j \cap (\sum_{i\neq j} M_i),$$

then m_j is identified by the isomorphism with the sequence having only one non-zero entry equal to m_j, and the sequence identified with $\sum_{i\neq j} m_i$ has the entry corresponding to j equal to zero, but by assumption both must be equal, so $m_j = 0$, as we wanted. ■

This result shows us that if each element of M can be uniquely written as a finite sum of elements from the family of submodules, we can naturally identify M as an external direct sum of R-modules. In this case, we write[3]

$$M = \bigoplus_{i\in\mathcal{I}} M_i,$$

and such a decomposition is called the *internal direct sum*, or simply the direct sum. In the case of modules, the external and internal direct sums are isomorphic, so we can identify elements of a direct sum of submodules both as almost-zero sequences and as finite sums. We can also observe that if L is a left ideal of R and $M = \bigoplus_{i\in\mathcal{I}} M_i$, then $LM = \bigoplus_{i\in\mathcal{I}} LM_i$, since certainly $LM = \sum_{i\in\mathcal{I}} LM_i$, and each $LM_i \subseteq M_i$, hence $LM_i \cap (\sum_{j\neq i} LM_j) = \{0\}$.

If R is a ring, then its left ideals are precisely its submodules, so if R can be written as a sum of submodules

$$R = \sum_{i\in\mathcal{I}} L_i,$$

where each L_i is a left ideal, and if it holds that any element in R can be uniquely written as a finite sum of elements in some of the left ideals L_i, we can write

$$R = \bigoplus_{i\in\mathcal{I}} L_i$$

as a direct sum of submodules, and this sum can be naturally identified with the external direct sum of the R-modules L_i.

The following result relates homomorphisms between modules expressed as direct sums and direct products, and will be of great interest for future proofs:

Proposition 1.3.8. Let R be a ring, $\{M_i\}_{i\in\mathcal{I}}$ a family of R-modules indexed by the set $\mathcal{I}$, and N an R-module. Then the following isomorphisms of abelian groups hold:

$$\mathrm{Hom}_R(\bigoplus_{i\in\mathcal{I}} M_i, N) \cong \prod_{i\in\mathcal{I}} \mathrm{Hom}_R(M_i, N);$$
$$\mathrm{Hom}_R(N, \prod_{i\in\mathcal{I}} M_i) \cong \prod_{i\in\mathcal{I}} \mathrm{Hom}_R(N, M_i).$$

In particular,

$$\mathrm{End}_R(\bigoplus_{i\in\mathcal{I}} M_i) \cong \prod_{i,j\in\mathcal{I}} \mathrm{Hom}_R(M_i, M_j)$$

[3]We will abuse notation and use the same symbol to denote both internal and external direct sums of modules, and when necessary, we will make explicit which sum we are referring to.

is an isomorphism of abelian groups, and if $\mathrm{Hom}_R(M_i, M_j) = \{0\}$ for $i \neq j$, we obtain that

$$\mathrm{End}_R(\bigoplus_{i\in\mathcal{I}} M_i) \cong \prod_{i\in\mathcal{I}} \mathrm{End}_R(M_i)$$

is an isomorphism of rings, where the right-hand side is a direct product of rings.

Proof. To prove the first isomorphism, let f be an R-homomorphism between $\bigoplus_{i\in\mathcal{I}} M_i$ and N, and consider the following map

$$\begin{aligned}\varphi : \mathrm{Hom}_R(\bigoplus_{i\in\mathcal{I}} M_i, N) &\mapsto \prod_{i\in\mathcal{I}} \mathrm{Hom}_R(M_i, N),\\ \varphi(f) &= (f \circ \iota_i)_{i\in\mathcal{I}},\end{aligned}$$

where ι_i denotes the canonical inclusion that maps an element m_i of M_i to the sequence with only the i-th entry non-zero and equal to m_i – note that $f \circ \iota_i$ can be viewed as the restriction of f to M_i, meaning φ simply sends f to the sequence of its restrictions to the respective components of the direct sum. From this definition, it follows that φ is a homomorphism of abelian groups, so we now check that it is bijective. If $\varphi(f) = 0$, it means that for every i in $\mathcal{I}$, and for any m_i in M_i, we have

$$(f \circ \iota_i)(m_i) = 0.$$

On the other hand, if $(m_i)_{i\in\mathcal{I}}$ is an element of $\bigoplus_{i\in\mathcal{I}} M_i$, we can write it as

$$(m_i)_{i\in\mathcal{I}} = \sum_{i\in\mathcal{I}} \iota_i(m_i),$$

where the sum on the right is finite because the sequence is almost zero, so

$$\begin{aligned}f((m_i)_{i\in\mathcal{I}}) &= f(\sum_{i\in\mathcal{I}} \iota_i(m_i))\\ &= \sum_{i\in\mathcal{I}} (f \circ \iota_i)(m_i) = 0,\end{aligned}$$

so f is identically zero, and thus φ is injective. Now, if $(f_i)_{i\in\mathcal{I}}$ is an element of $\prod_{i\in\mathcal{I}} \mathrm{Hom}_R(M_i, N)$, we define an R-homomorphism f such that

$$f((m_i)_{i\in\mathcal{I}}) = \sum_{i\in\mathcal{I}} f_i(m_i) \in N,$$

for any sequence $(m_i)_{i\in\mathcal{I}}$ in $\bigoplus_{i\in\mathcal{I}} M_i$, and since each sequence is almost zero, the sum on the right-hand side of the equation is finite. From this, we observe that $(f \circ \iota_i)(m_i) = f_i(m_i)$, so $\varphi(f) = (f_i)_{i\in\mathcal{I}}$, and thus φ is surjective.

To prove the second isomorphism, let $f \in \mathrm{Hom}_R(N, \prod_{i\in\mathcal{I}} M_i)$, and define

$$\begin{aligned}\psi : \mathrm{Hom}_R(N, \prod_{i\in\mathcal{I}} M_i) &\mapsto \prod_{i\in\mathcal{I}} \mathrm{Hom}_R(N, M_i),\\ \psi(f) &= (\pi_i \circ f)_{i\in\mathcal{I}},\end{aligned}$$

where π_i denotes the canonical projection of the direct product onto its i-th component – note that ψ is simply the map that sends f to the sequence of its projections onto the respective components of the direct product. Again, the definition makes it clear that ψ is a homomorphism of abelian groups, so we need to check that it is a bijection. If $\psi(f) = 0$, it means that for every $i \in \mathcal{I}$ and for every $n \in N$, we have

$$(\pi_i \circ f)(n) = 0,$$

but on the other hand, we can write

$$f(n) = ((\pi_i \circ f)(n))_{i\in\mathcal{I}},$$

so f is identically zero, and thus ψ is injective. Now, if $(f_i)_{i\in\mathcal{I}}$ is an element of $\prod_{i\in\mathcal{I}} \mathrm{Hom}_R(N, M_i)$, we define the R-homomorphism f given by

$$f(n) = (f_i(n))_{i\in\mathcal{I}} \in \prod_{i\in\mathcal{I}} M_i,$$

and then it follows that $\pi_i \circ f = f_i$, so

$$\psi(f) = ((\pi_i \circ f))_{i\in\mathcal{I}} = (f_i)_{i\in\mathcal{I}},$$

which allows us to conclude that ψ is surjective.

The third isomorphism stated can be obtained by noting that, from the previous considerations, we have

$$\mathrm{Hom}_R(\bigoplus_{i\in\mathcal{I}} M_i, \bigoplus_{i\in\mathcal{I}} M_i) \overset{\varphi}{\cong} \prod_{i\in\mathcal{I}} \mathrm{Hom}_R(M_i, \bigoplus_{i\in\mathcal{I}} M_i) \overset{\psi}{\cong} \prod_{i,j\in\mathcal{I}} \mathrm{Hom}_R(M_i, M_j),$$

where the elements of $\prod_{i,j\in\mathcal{I}} \mathrm{Hom}_R(M_i, M_j)$ are of the form

$$(\pi_j \circ f \circ \iota_i)_{i,j\in\mathcal{I}} = ((\pi_j \circ f \circ \iota_i)_{i\in\mathcal{I}})_{j\in\mathcal{I}},$$

since the composition of compatible additive group isomorphisms is an isomorphism, so $\rho = \psi \circ \phi$ is the desired isomorphism between $\mathrm{End}_R(\bigoplus_{i\in\mathcal{I}} M_i)$ and $\prod_{i,j\in\mathcal{I}} \mathrm{Hom}_R(M_i, M_j)$. Now note that if $\mathrm{Hom}_R(M_i, M_j) = \{0\}$ when $i \neq j$, then we have

$$(\pi_j \circ f \circ \iota_i)_{i,j\in\mathcal{I}} = (\pi_i \circ f \circ \iota_i)_{i\in\mathcal{I}},$$

since $\pi_j \circ f \circ \iota_i = 0$ if $i \neq j$, so $\prod_{i,j\in\mathcal{I}} \mathrm{Hom}_R(M_i, M_j) = \prod_{i\in\mathcal{I}} \mathrm{End}_R(M_i)$, and from this we get

$$\mathrm{End}_R(\bigoplus_{i\in\mathcal{I}} M_i) \overset{\rho}{\cong} \prod_{i\in\mathcal{I}} \mathrm{End}_R(M_i),$$

where $\rho(f) = (\pi_i \circ f \circ \iota_i)_{i\in\mathcal{I}}$, and the isomorphisms are additive group isomorphisms, so to prove the final statement, it is enough to check that they are also ring homomorphisms. The identity element in $\mathrm{End}_R(\bigoplus_{i\in\mathcal{I}} M_i)$ is the identity function id, so

$$\rho(\mathrm{id}) = (\pi_i \circ \mathrm{id} \circ \iota_i)_{i\in\mathcal{I}} = (\mathrm{id}_i)_{i\in\mathcal{I}},$$

since $\pi_i \circ \iota_i = \mathrm{id}_i$ – where id_i is the identity in M_i –, so the image of the identity in the domain is the identity in the codomain. If f, g are endomorphisms in the domain, then

$$\begin{aligned}
\rho(f \circ g) &= (\pi_i \circ f \circ \iota_i)_{i\in\mathcal{I}} \\
&= (\pi_i \circ f \circ \iota_i \circ \pi_i \circ g \circ \iota_i)_{i\in\mathcal{I}} \\
&= ((\pi_i \circ f \circ \iota_i) \circ (\pi_i \circ g \circ \iota_i))_{i\in\mathcal{I}} \\
&= \rho(f) \circ \rho(g),
\end{aligned}$$

so the isomorphism is indeed a ring homomorphism, as desired. ■

Note that if the domain of an R-homomorphism f is a direct sum of modules, then f is uniquely determined by the restrictions $(f \circ \iota_i)_{i\in\mathcal{I}}$ of f on the components of the domain. Similarly, if its codomain is a direct product of modules, then f is uniquely determined by the functions $(\pi_i \circ f)_{i\in\mathcal{I}}$, and if both domain and codomain are direct sums, it follows that the elements of the form $(\pi_i \circ f \circ \iota_j)_{i,j}$ determine f uniquely. When the direct sums are finite, the isomorphism allows us to uniquely identify f with a matrix whose entry in position ij is given by $(\pi_i \circ f \circ \iota_j)_{i,j}$. It is also worth noting that if R is a commutative ring, then the isomorphisms above will also be R-module isomorphisms.

The ring of R-endomorphisms of a ring R is also directly related to its opposite ring.

Proposition 1.3.9. The rings $\mathrm{End}_R(R)$ and R^{op} are isomorphic.

Proof. Define the following homomorphism of abelian groups:

$$\varphi : R^{\text{op}} \mapsto \text{End}_R(R);$$
$$\varphi(a)(b) = b \cdot a.$$

Note that $\varphi(a * b) = \varphi(b \cdot a)$, so for any $x \in R$, we have

$$\varphi(b \cdot a)(x) = x \cdot b \cdot a = \varphi(a) \circ \varphi(b)(x),$$

and also that $\varphi(1)$ is the identity function in $\text{End}_R(R)$, so the homomorphism is a ring homomorphism. If $\varphi(a) = 0$, we have that for any $b \in R$, $b \cdot a = 0$, and since 1 is also an element of R, this implies that $a = 0$, so φ is injective. Given any R-endomorphism f of R, we have that for all $x \in R$

$$f(x) = f(x \cdot 1) = x \cdot f(1) = \varphi(f(1))(x),$$

implying that φ is a ring isomorphism. ■

1.3.4 Simple Modules

Now we will prove one of the most important results about simple R-modules, known as Schur's Lemma.

Lemma 1.3.10 (Schur's Lemma)**.** Let M and N be non-zero simple R-modules, and let $\varphi : M \mapsto N$ be an R-homomorphism. Then φ is either identically zero or an isomorphism. In particular, the ring $\text{End}_R(M)$ is a division ring.

Proof. We know that the sets $\ker(\varphi)$ and $\text{Im}(\varphi)$ are submodules of M and N, respectively. Since M and N are simple, it follows that $\ker(\varphi) = \{0\}$ or $\ker(\varphi) = M$, and that $\text{Im}(\varphi) = \{0\}$ or $\text{Im}(\varphi) = N$. If $\ker(\varphi) = \{0\}$, by Theorem 1.3.3, we have that $M \cong_R \text{Im}(\varphi)$, and therefore $M \cong_R N$, since M and N are non-zero. Otherwise, φ is identically zero.

Now, consider the set $\text{End}_R(M)$. We know that it is indeed a ring with addition and composition of functions, and with the identity given by the identity function. It suffices to show that every non-zero element is invertible. Given any element f of the ring of endomorphisms, since M is simple, there are two options: $\ker(f) = \{0\}$, and in this case, by the same argument used earlier, it follows that $\text{Im}(f) = M$, or $\ker(f) = M$, and in this case, $f = 0$. Therefore, if f is not identically zero, then it is a bijection, and thus invertible, so $\text{End}_R(M)$ is a division ring. ■

Schur's Lemma allows us to characterize simple modules over any ring.

Theorem 1.3.11. Let M be an R-module. The following are equivalent:

(1) M is simple;

(2) M is cyclic, and every non-zero element of M is a generator;

(3) M is isomorphic as an R-module to R/L, where L is a maximal left ideal of R.

Proof.

(1)⇒(2) Assume M is simple, and note that for any non-zero element m of M, Rm is an R-submodule containing at least 0 and m, so $Rm \neq \{0\}$, and by simplicity, we have $Rm = M$.

(2)⇒(3) Assume that M is cyclic and that every non-zero element is a generator. Again, consider a non-zero element m of M, and define

$$\varphi : R \mapsto M,$$
$$\varphi(r) = rm,$$

and note that this function is clearly an R-epimorphism, so by Theorem 1.3.3, we have

$$R/\ker(\varphi) \cong_R M.$$

Now note that for any $r_1, r_2 \in \ker(\varphi)$,

$$\varphi(r_1 - r_2) = \varphi(r_1) - \varphi(r_2) = 0,$$

so $r_1 - r_2 \in \ker(\varphi)$, and also for any $r \in R$,

$$\varphi(rr_1) = rr_1m = r(r_1m) = 0,$$

so $rr_1 \in \ker(\varphi)$, and thus $\ker(\varphi)$ is a left ideal of R. If J is a left ideal of R such that $\ker(\varphi) \subsetneq J$, then there exists $r \in J \setminus \ker(\varphi)$, so $\varphi(r) = rm \neq 0$, and by simplicity, we would have $M = R(rm)$. Thus, the restriction $\varphi|_J = \varphi \circ \iota$, where ι is the inclusion of J in R, is surjective, and therefore

$$J/\ker(\varphi) \cong_R M \cong_R R/\ker(\varphi),$$

so $J = R$, implying that $\ker(\varphi)$ is a maximal left ideal.

(3)⇒(1) Assume that M is isomorphic to R/L, where L is a maximal left ideal of R. If N is an R-submodule of M, then the isomorphism shows that N is isomorphic to some submodule of R/L, so by Theorem 1.3.4, there exists some R-submodule J of R containing L such that

$$N \cong_R J/L.$$

On the other hand, since R is a ring, we know that its R-submodules are precisely its left ideals, so J is a left ideal containing the maximal left ideal L, implying that $J = R$ or $J = L$. In both cases, this means that N is isomorphic to R/L or to $\{0\}$, and by the arbitrariness of N, we conclude that M is simple. ■

The previous result allows us to identify maximal left ideals of a ring R with the possible simple R-modules. In particular, since every simple R-submodule of R is a minimal left ideal, we have that each minimal left ideal is isomorphic as an R-module to the quotient of R by some maximal left ideal. However, it is not true that any simple R-module (not necessarily contained in R) is isomorphic to some minimal ideal of R, simply because not all rings have minimal left ideals, but all have maximal left ideals. An example of this is the ring $\mathbb{Z}$, which does not have minimal ideals because its ideals are of the form $n\mathbb{Z}$ for some integer n, but there exist simple $\mathbb{Z}$-modules, e.g., $\mathbb{Z}_p$ with p prime is simple because it is isomorphic to $\mathbb{Z}/p\mathbb{Z}$, and $p\mathbb{Z}$ is a maximal ideal.

An important set in the study of modules is the so-called *annihilator*. Formally, if M is a left R-module and $x \in M$, then

$$\mathrm{Ann}_R(x) := \{a \in R | ax = 0\} \subseteq R,$$

and it is immediate to check that $\mathrm{Ann}_R(x)$ is a left ideal of R. If $N \subseteq M$ is a subset of M, then the annihilator $\mathrm{Ann}_R(N)$ of N is simply the intersection of the annihilators of its elements, and it is also immediate to check that $\mathrm{Ann}_R(M)$ is a two-sided ideal of R. We say that an R-module M is *faithful* if $\mathrm{Ann}_R(M) = \{0\}$.

Now we can describe in more detail the minimal left ideals of a ring.

Lemma 1.3.12 (Brauer's Lemma)**.** Let L be a minimal left ideal of a ring R. Then $L^{(2)} = \{0\}$ or there exists an idempotent element $e \in L$ such that $L = Re$, and in this case, eRe is a division ring.

Proof. Assume that $L^{(2)} \neq \{0\}$, so there exists a non-zero element $a \in L$ such that $La \neq \{0\}$. However, La is a non-trivial left ideal of L, so by the simplicity of L we have $La = L$, and thus we can find some element $e \in L$ such that $ea = a$. Now we note that

$$(e^2 - e)a = ea - a = 0,$$

so $e^2 - e$ belongs to the annihilator $\text{Ann}_L(a)$ of a in L, but on the other hand, we know that this annihilator is a left ideal of L, which must be proper since $ea = a$. Therefore, by the simplicity of L, we have $\text{Ann}_L(a) = \{0\}$, implying that $e^2 = e$. Since L is simple, it follows from Theorem 1.3.11 that $L = Re$.

To prove the second statement, we first note that eRe is certainly a ring with unity given by e, so it is enough to show that it is a division ring. We will construct an isomorphism from eRe to the ring $\text{End}_R(Re)$, and this together with Schur's Lemma will imply the desired result. First, note that if φ is an R-endomorphism of Re, then for any $a \in R$ we have

$$\varphi(ae) = \varphi(a(ee)) = a\varphi(e),$$

i.e., φ is completely determined by $\varphi(e)$. Furthermore, if $\varphi, \varphi' \in \text{End}_R(Re)$, and if $\varphi'(e) = ae$, then

$$\varphi(\varphi'(e)) = \varphi(ae) = a\varphi(e) = \varphi'(e)\varphi(e),$$

since $e\varphi(e) = \varphi(e)$, we can consider the following map:

$$\begin{aligned} \psi : \text{End}_R(Re) &\mapsto eRe, \\ \psi(\varphi) &= e\varphi(e). \end{aligned}$$

The map is clearly a homomorphism of abelian groups, and from the previous considerations, we have that

$$\psi(\varphi \circ \varphi') = e\varphi(\varphi'(e)) = e\varphi'(e)\varphi(e) = (e\varphi'(e))(e\varphi'(e)) = \psi(\varphi)\psi(\varphi'),$$

so ψ is a ring homomorphism. It will be surjective because for any $a \in R$ we can define $\varphi \in \text{End}_R(Re)$ from $\varphi(e) = a$, and if $\psi(\varphi) = 0$, it means that $\varphi(e) = e\varphi(e) = 0$, so φ is identically zero. This shows that ψ is an isomorphism, and since Re is simple, it follows from Schur's Lemma that $\text{End}_R(Re)$ is a division ring, so eRe is also a division ring. ■

The previous results allow us to conclude important facts about the ring of matrices with entries from some division ring.

Example 1.3.13 (Semisimplicity of $M_n(D)$)**.** Consider the ring of matrices $M_n(D)$ with entries in a division ring D and with $n > 1$, and note that we can write this ring as

$$M_n(D) = C_1 \oplus ... \oplus C_n,$$

where each C_i is a submodule of the matrices with the i-th column non-zero, given by

$$C_i = \{ \begin{pmatrix} 0 & ... & a_{1i} & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & ... & a_{ni} & ... & 0 \end{pmatrix} | a_{1i}, ..., a_{ni} \in D \}.$$

It is immediate to notice that the sets C_i are indeed left ideals of $M_n(D)$, each naturally isomorphic to D^n as a $M_n(D)$-module, however we can also observe that these ideals are minimal. Indeed, take any left ideal L contained in some C_i, and assume that there exists a non-zero element y in L, i.e., there exists an index j such that the entry $y_{ji} \neq 0$ in the matrix y. Now take any element $x \in C_i$, and note that the matrix

$$M = \begin{pmatrix} 0 & ... & x_{1i}y_{ji}^{-1} & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & ... & x_{ni}y_{ji}^{-1} & ... & 0 \end{pmatrix}$$

is such that $My = x$, and since L is a left ideal, it follows that $My \in L$, so $x \in L$, which implies that $C_i = L$, and therefore C_i is a minimal left ideal. This allows us to conclude that the set of matrices can be written as a direct sum of minimal left ideals, something we will define in the future as semisimplicity. Since $C_i^{(2)} \neq \{0\}$, Brauer's Lemma guarantees that these ideals will be of the form $M_n(D)E$, where E is some idempotent matrix, and in the case, it suffices to take for any C_i the matrix E_{ii}, which is certainly idempotent and belongs to C_i, so indeed we have

$$M_n(D) = M_n(D)E_{11} \oplus ... \oplus M_n(D)E_{nn}.$$

1.3.5 Vector Spaces and Zorn's Lemma

The previous example also illustrates an interesting fact: a ring R being simple does not necessarily imply that it is simple as an R-module. In fact, we saw earlier that $M_n(D)$ is a simple ring, but the previous example shows that, when viewed as a module over itself, this ring is not simple for the cases where $n > 1$, as it can be written as a direct sum of n distinct simple submodules. The decomposition in this example is relatively simple, yet it captures the main motivation behind the results in the upcoming chapters: to describe rings as a direct sum of simple objects in an explicit manner. It is also possible to characterize all the simple submodules of the matrix ring, something that will be particularly useful for applications of Wedderburn's Theorem.

Example 1.3.14 (Uniqueness of Simple Submodules of $M_n(D)$)**.** Let $R = M_n(D)$ be the ring of matrices over a division ring D, and consider a simple R-module M. By Theorem 1.3.11, it follows that $M = RX$, where X is a non-zero matrix in M, and since $RX \neq 0$, there certainly exists some non-zero matrix A in R such that $AX \neq 0$. Fix a column j of A with some non-zero entry, and consider the map φ from M to D^n that sends an element BX of M to the j-th column of B, i.e., $\varphi(BX) = Be_j$, where e_j is the vector with one in the j-th entry and zero elsewhere. It follows that this map is a non-zero R-homomorphism, and from the previous example, we saw that D^n is a simple R-module, so by Schur's Lemma, φ must be an R-isomorphism, thus $M \cong_R D^n$. This shows us that, up to isomorphism, D^n is the only simple $M_n(D)$-module.

We mentioned that modules generalize the notion of a vector space, as a vector space is nothing more than a module over a field, but this generalization causes these objects to lose several useful properties. Below we list some properties that are generally false for modules over rings but are always true for vector spaces over fields:

- Every linearly independent set can be extended to a basis;
- Every generating set contains a basis;
- Every element of a linearly independent set is a linear combination of the others;
- Every submodule of a free module is free.

We will now state without proof an important result for future sections. To do this, we remind the reader that given a set equipped with a binary relation between its elements, we say it is a partially ordered set – also called a *poset* – if the relation is reflexive, antisymmetric, and transitive, and we say that a subset is a chain if it is totally ordered, i.e., every pair of elements is comparable.

Lemma 1.3.15 (Zorn's Lemma)**.** Let $(X, \leq)$ be a poset where every chain $C \subseteq X$ has an upper bound in X, that is, an element x in X such that for every y in C, $y \leq x$. Then X has a maximal element.

Zorn's Lemma is an extremely powerful tool for proving various results in linear algebra, and it will be particularly useful for demonstrating basic properties of semisimple modules and rings. Below, we provide an example of using Zorn's Lemma to demonstrate that every ideal in a ring is contained in some maximal ideal.

Example 1.3.16 (Existence of Maximal Ideals)**.** Let R be a ring, and fix a left ideal L. Define the set

$$\mathcal{F} = \{J \subseteq R \mid L \subseteq J,\ J \text{ is a left ideal of } R\}.$$

The set in question is a poset with the inclusion relation $\subseteq$, so we take a chain C in $\mathcal{F}$. We can define $\tilde{J} = \bigcup_{J \in C} J$, and we claim that this element is an upper bound for C. In fact, first note that $L \subseteq \tilde{J}$ by definition, and if $a, b \in \tilde{J}$, there exist $J_1, J_2 \in C$ such that $a \in J_1, b \in J_2$, and assuming without loss of generality that $J_1 \subseteq J_2$, we have $a, b \in J_2$, so $a + b \in J_2$, implying that $\tilde{J}$ is closed under addition. Now, if $a \in R$ and $b \in \tilde{J}$, a similar argument shows that $ab \in \tilde{J}$, so $\tilde{J}$ belongs to $\mathcal{F}$, and then by Zorn's Lemma, it follows that there exists a maximal element in $\mathcal{F}$, which in this case will be exactly a maximal left ideal of R containing L. A similar argument can be made for right ideals and two-sided ideals.

1.4 Algebras

The main object of study in this work is what are called algebras over fields, which are nothing more than vector spaces with some notion of bilinear product between their elements. Below is its formal definition.

Definition 1.4.1 (Algebra)**.** Let $(\mathcal{A}, +)$ be an abelian group and $\mathbb{F}$ a field. We say that $\mathcal{A}$ is an $\mathbb{F}$-algebra if there exist operations $\cdot : \mathbb{F} \times \mathcal{A} \mapsto \mathcal{A}$ and $* : \mathcal{A} \times \mathcal{A} \mapsto \mathcal{A}$ such that:

(1) $(\mathcal{A}, +, \cdot)$ is a vector space over $\mathbb{F}$;

(2) $(\mathcal{A}, +, *)$ is a ring[4];

(3) For any α in $\mathbb{F}$ and A, B in $\mathcal{A}$: $\alpha \cdot A * B = (\alpha A) * B = A * (\alpha B)$.

Items (1) and (2) tell us that an algebra is nothing more than a vector space with respect to a field $\mathbb{F}$ that is also a ring, and (3) tells us that the multiplication between elements of the ring is compatible with scalar multiplication from $\mathbb{F}$, and therefore all the constructions and results seen so far about rings and modules also apply.

The set $M_n(\mathbb{C})$ of $n \times n$ matrices with complex entries is a ring with respect to the matrix addition and multiplication operations, and its unit is the identity matrix. Furthermore, this set also has the structure of a $\mathbb{C}$-vector space with scalar multiplication by elements of $\mathbb{C}$, so it is a $\mathbb{C}$-algebra. We can also consider the term-by-term multiplication operation of matrices, known as the Schur product, defined as $(A \circ B)_{ij} = A_{ij}B_{ij}$ for matrices of the same size, and we can observe that $M_n(\mathbb{C})$ is also a $\mathbb{C}$-algebra with respect to the Schur product, with the unit given by the matrix with all elements equal to one.

Given an algebra $\mathcal{A}$ over a field $\mathbb{F}$, we will denote its unit by the symbol E. A subset $\mathcal{B}$ of $\mathcal{A}$ that is simultaneously a subspace and a subring is called a *subalgebra*, meaning that $\mathcal{B}$ is a subspace of $\mathbb{F}$ that is closed under the matrix product and contains the unit E of $\mathcal{A}$. The center $Z(\mathcal{A})$ of an algebra is always a subalgebra, simply due to the compatibility between the algebra's operations: if A commutes with all elements of $\mathcal{A}$, then certainly αA also commutes for any $\alpha \in \mathbb{F}$. A function φ between $\mathbb{F}$-algebras is called a *homomorphism* if it is simultaneously a homomorphism of $\mathbb{F}$-vector spaces and a homomorphism of rings.

We say that a set M is an $\mathcal{A}$-module if M is a module over the ring $\mathcal{A}$, and in this case, we can provide a structure of $\mathbb{F}$-vector space to M by defining the product

$$\alpha m := (\alpha E)m,$$

for any $\alpha \in \mathbb{F}, m \in M$, and it follows that

$$(\alpha X)m = \alpha(Xm) = X(\alpha m)$$

for any $X \in \mathcal{A}$, and therefore an $\mathcal{A}$-homomorphism is also an $\mathbb{F}$-homomorphism.

The notions of direct sums and direct products naturally translate into the context of algebras, but in this case, we must make an additional effort to clarify which type of sum or product we are referring to. In this work, a direct sum of algebras over a field $\mathbb{F}$ will always refer to a direct sum of $\mathbb{F}$-vector spaces, that is, a direct sum of $\mathbb{F}$-modules, and a direct product of algebras will refer to an external direct product of rings, since each algebra is also a ring.

An algebra $\mathcal{A}$ over a field $\mathbb{F}$ that is an $\mathbb{F}$-vector space of finite dimension d is said to be a *finite-dimensional algebra*. In this case, given that we fixed a basis $\{A_1, ..., A_d\}$ for $\mathcal{A}$, we have

$$A_i A_j = \sum_{k=1}^{d} c_{ij}^k A_k,$$

[4]Many other works related to the study of algebras also study algebras that do not necessarily have a unit, and in this case, the algebras described here would be called *unital*.

for any $i, j \in [d] := \{1, ..., d\}$, where $c_{ij}^k \in \mathbb{F}$. Therefore, note that given elements $A, B \in \mathcal{A}$ written as $A = \sum_i \alpha_i A_i$ and $B = \sum_j \beta_j A_j$, we have

$$\begin{aligned} AB &= (\sum_i \alpha_i A_i)(\sum_j \beta_j A_j) \\ &= \sum_{i,j} \alpha_i \beta_j A_i A_j \\ &= \sum_{ijk} \alpha_i \beta_j c_{ij}^k A_k, \end{aligned}$$

that is, if we fix a basis, the constants $\{c_{ij}^k\}_{ijk}$ completely determine the algebra $\mathcal{A}$, and they are usually called the *structural constants* of $\mathcal{A}$.

2 Semisimplicity

Now we will focus on studying what are called semisimple rings and modules. This concept is of great importance in various areas of mathematics and physics, such as in the study of group representations and Lie algebras, and the structure of semisimple algebras will be useful in proving several results for graphs and optimization programs. In this chapter, we will present the theory of semisimplicity in its most general form, and later focus on applications in the following chapters. But before that, we will briefly discuss one of the motivations behind our study.

One of the most famous results in classical linear algebra is the Primary Decomposition Theorem, which tells us that given an endomorphism of vector spaces f on a space V over a field $\mathbb{F}$, it is always possible to decompose the space as

$$V = W_1 \oplus \ldots \oplus W_k,$$

where each W_i is a subspace invariant under f, and in particular, it will be a generalized eigenspace. That is, if we write the minimal polynomial of f as

$$m_f(t) = p_1(t)^{r_1} \cdot \ldots \cdot p_k^{r_k}(t),$$

with each p_i irreducible, then $W_i := \ker(p_i^{r_i}(f))$. A decomposition of the space into a direct sum of f-invariant subspaces naturally gives us a block-diagonalization of f: simply choose a basis for each W_i and then take the union of such bases. In the context of matrices with complex entries, the theorem guarantees that it is always possible to find a block-diagonal basis for a given matrix, and much of basic linear algebra is dedicated to understanding the form of these blocks. For example, over the complex numbers, it will always be possible to triangularize the blocks of the primary decomposition of a matrix, so that each block has a diagonal component and a nilpotent component, and with a little more effort, we can obtain the famous Jordan canonical form. In the case of diagonalizable matrices, their primary decomposition is exactly their spectral decomposition, meaning that each block will have size 1, and its only entry will be the eigenvalue associated with the respective eigenspace.

We can then ask ourselves how to generalize these results: if we now have a set of matrices, is it always possible to find a basis that simultaneously puts them in a block-diagonal form? We will see that the answer to this question is closely connected with the notion of semisimplicity.

2.1 Semisimple Modules

Given any R-module M, we say that M is *semisimple* if it can be written as a direct sum of simple R-modules. We are interested in characterizing semisimple modules, and to do this, we will first prove the following auxiliary result:

Lemma 2.1.1. Let M be an R-module such that every submodule N of M is a direct summand, that is, there exists a submodule N' of M such that $M = N \oplus N'$. Then every non-zero submodule of M has a simple submodule.

Proof. Let M be an R-module as described in the statement, and let N be a non-zero submodule. The idea of this proof is to use the existence of a maximal left ideal of R to derive a contradiction with the non-existence of a simple submodule of N. To do this, first fix a non-zero element $x \in N$, and consider the following map:

$$\varphi : R \mapsto Rx;$$
$$\varphi(a) = ax.$$

It is immediate to check that φ is an epimorphism of R-modules, and it is also worth noting that the kernel of φ is precisely the annihilator $\mathrm{Ann}_R(x)$ of x in R, and this is properly contained in R since $x \neq 0$. Since $\mathrm{Ann}_R(x)$ is a proper left ideal of R, by Example 1.3.16, there exists a maximal left ideal L containing it. We then note that since L is maximal in R, by the Correspondence Theorem 1.3.4, $L/\mathrm{Ann}_R(x)$ is maximal

in $R/\mathrm{Ann}_R(x)$, and since $R/\mathrm{Ann}_R(x)$ is isomorphic as an R-module to Rx, it follows that the image Lx of $L/\mathrm{Ann}_R(x)$ under the isomorphism is a maximal submodule of Rx.

Now we will use the hypothesis to find a direct sum decomposition of Rx in terms of Lx. By hypothesis, we have $M = Lx \oplus L'$, where L' is some submodule of M, so for any $a \in R$, there exist unique $b \in L$, $y \in L'$ such that

$$ax = bx + y,$$

thus y can be uniquely written as

$$y = ax - bx \in L' \cap Rx,$$

implying that $Rx = Lx \oplus (Rx \cap L')$. We can now use the maximality of Lx to show that $Rx \cap L'$ is simple. Indeed, assume it is not, then there exists a proper non-trivial submodule of $Rx \cap L'$, and since this is also a submodule of Rx, it follows that it must be of the form Sx, for some non-trivial left ideal S of R. Consider the following chain of inclusions:

$$Lx \subsetneq Lx + Sx \subsetneq Lx + (Rx \cap L') = Rx,$$

and note that this implies $Lx + Sx$ is a proper submodule of Rx that properly contains the maximal submodule Lx, which clearly contradicts the maximality of Lx. Thus, we conclude that $Rx \cap L'$ is simple, as desired. ■

The previous result, together with Zorn's Lemma, allows us to prove the following characterization of semisimple modules:

Theorem 2.1.2. Let M be an R-module. The following are equivalent:

(1) M is a direct sum of simple R-modules;

(2) M is semisimple;

(3) Every R-submodule of M is a direct summand.

Proof.

(1)⇒(2) Assume that $M = \sum_{i\in\mathcal{I}} M_i$, where each M_i is a simple R-module, and consider the set

$$\mathcal{F} = \{\mathcal{J} \subseteq \mathcal{I} \mid \sum_{j\in\mathcal{J}} M_j \text{ is a direct sum}\}.$$

Note that $\mathcal{F} \neq \emptyset$, since each $M_i \in \mathcal{F}$, and that $(\mathcal{F}, \subseteq)$ is a partially ordered set (poset). Let $C \subseteq \mathcal{F}$ be a chain, and note that $\bigcup_{\mathcal{J}\in C} \mathcal{J} \in \mathcal{F}$ is an upper bound for C, that is, every chain has an upper bound in $\mathcal{F}$. By Zorn's Lemma, there exists a maximal $\mathcal{J}_m \in \mathcal{F}$, and note that given any M_i, the intersection $M_i \cap \sum_{j\in\mathcal{J}_m} M_j$ is a submodule of M_i, so

$$M_i \cap \sum_{j\in\mathcal{J}_m} M_j = M_i \qquad \text{or} \qquad M_i \cap \sum_{j\in\mathcal{J}_m} M_j = \{0\},$$

because M_i is simple. Assume that there exists some $i \in \mathcal{I}$ such that the second case occurs, and note then that $\mathcal{J}_m \cup \{i\} \in \mathcal{F}$ is a set of indices that results in a direct sum and strictly contains the maximal set $\mathcal{J}_m$, which contradicts the definition of this set. Thus, every M_i is contained in $\sum_{j\in\mathcal{J}_m} M_j$, and therefore $M = \sum_{j\in\mathcal{J}_m} M_j$, that is, M is semisimple.

(2)⇒(3) Assume that $M = \bigoplus_{i\in\mathcal{I}} M_i$, where each M_i is a simple submodule. Let N be a submodule of M, and consider the set

$$\mathcal{F} = \{\mathcal{J} \subseteq \mathcal{I} \mid N + \sum_{j\in\mathcal{J}} M_j \text{ is a direct sum}\}.$$

Analogously to the previous step, it is immediate to observe that $(\mathcal{F}, \subseteq)$ is a poset. Note that $\mathcal{F}$ is non-empty, because for each M_i, either $M_i \cap N = \{0\}$ or $M_i \cap N = M_i$ (since M_i is simple), and if for all i, we have

$M_i \cap N = M_i$, then $M = N$, and there is nothing to prove. Thus, we can assume that there exists at least one i such that $M_i \cap N = \{0\}$, and therefore $N + M_i$ is a direct sum. Let $C \subseteq \mathcal{F}$ be a chain, and note that $\bigcup_{\mathcal{J} \in C} \mathcal{J} \in \mathcal{F}$ is an upper bound for C in $\mathcal{F}$, and then by Zorn's Lemma, there exists a maximal $\mathcal{J}_m \in \mathcal{F}$ in $\mathcal{F}$. Note that if there exists M_i such that $M_i \cap (N + (\bigoplus_{j \in \mathcal{J}_m} M_j)) = \{0\}$, then $\mathcal{J}_m \cup \{i\} \in \mathcal{F}$, contradicting the maximality of $\mathcal{J}_m$ in $\mathcal{F}$. Therefore, all the M_i's are contained in the sum, so we have

$$M = N \oplus (\bigoplus_{j \in \mathcal{J}_m} M_j),$$

and thus N is a direct summand.

(3)⇒(1) Assume that every submodule of M is a direct summand, and consider N_s the submodule of M formed by the sum of all the simple submodules of M. By hypothesis, there exists a submodule N' such that

$$M = N_s \oplus N'.$$

If $N' \neq \{0\}$, Lemma 2.1.1 guarantees that there exists a simple submodule of N', which contradicts the definition of N_s, so $N' = \{0\}$, and therefore M is semisimple. ■

The previous proposition shows us that semisimple R-modules are precisely those that, in a sense, behave similarly to vector spaces: given any submodule, it is possible to find a complementary module. Semisimplicity also behaves well with respect to submodules and quotients, as we can verify below:

Corollary 2.1.3. Every submodule and every quotient of a semisimple R-module is semisimple.

Proof. Let $M = \bigoplus_{i \in \mathcal{I}} M_i$ be a semisimple R-module, with each M_i simple, and take any submodule N. Consider the R-epimorphism projection $\pi : M \mapsto M/N$, and note that for any $m \in M$, there exist unique m_{i_j} such that

$$m = \sum_j m_{i_j},$$

so $\pi(m) = \sum \pi(m_{i_j})$ uniquely, and hence $M/N = \sum_{i \in \mathcal{I}} \pi(M_i)$, which implies by the previous proposition that M/N is semisimple. From the previous proposition, we also know that there exists a submodule N' such that $M = N \oplus N'$, so it follows that $M/N' \cong N$, and thus N is isomorphic to a quotient, and therefore is semisimple as well. ■

Now we will formalize a way to identify endomorphisms of semisimple modules with matrices, which was briefly mentioned right after the proof of Proposition 1.3.8.

Proposition 2.1.4. Let M be a module over a ring R, and consider M^n, where n is a natural number. Then $\mathrm{End}_R(M^n)$ and $M_n(\mathrm{End}_R(M))$ are isomorphic rings.

Proof. Let $f \in \mathrm{End}_R(M^n)$, and let π_i, ι_j be the canonical projection and inclusion maps onto the i-th and j-th copies of M, respectively, where $\pi_i \circ f \circ \iota_j \in \mathrm{End}_R(M)$. It is worth noting that, although they are copies of M, the functions $\pi_i \circ f \circ \iota_j$ are generally not equal. Observe that if $m = (m_1, \ldots, m_n)$ is an element of M^n, then

$$f(m) = (\sum_{j=1}^{n} (\pi_1 \circ f \circ \iota_j)(m_j), \ldots, \sum_{j=1}^{n} (\pi_n \circ f \circ \iota_j)(m_j)),$$

so if we identify m with a column vector, we have

$$f(m) = \begin{pmatrix} \pi_1 \circ f \circ \iota_1 & \pi_1 \circ f \circ \iota_2 & \ldots & \pi_1 \circ f \circ \iota_n \\ \vdots & \vdots & \vdots & \vdots \\ \pi_n \circ f \circ \iota_1 & \pi_n \circ f \circ \iota_2 & \ldots & \pi_n \circ f \circ \iota_n \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}.$$

Therefore, if we consider the map

$$\varphi : \mathrm{End}_R(M^n) \mapsto M_n(\mathrm{End}_R(M)),$$
$$\varphi(f) = \begin{pmatrix} \pi_1 \circ f \circ \iota_1 & \pi_1 \circ f \circ \iota_2 & \dots & \pi_1 \circ f \circ \iota_n \\ \vdots & \vdots & \vdots & \vdots \\ \pi_n \circ f \circ \iota_1 & \pi_n \circ f \circ \iota_2 & \dots & \pi_n \circ f \circ \iota_n \end{pmatrix},$$

it is immediate to observe that it is a group homomorphism, and if $f, g \in \mathrm{End}_R(M^n)$, the previous observations guarantee that $\varphi(f \circ g) = \varphi(f) \cdot \varphi(g)$, and since $\varphi(\mathrm{id}) = I$, it follows that we have a ring homomorphism. If $\varphi(f) = 0$, as we saw in Proposition 1.3.8, f is uniquely determined by the entries of the matrix $\varphi(f)$, so $f = 0$. Given any matrix M in $M_n(\mathrm{End}_R(M))$, observe that there exists a unique R-homomorphism f such that $\pi_i \circ f \circ \iota_j = M_{ij}$, so $\varphi(f) = M$. Therefore, we conclude that φ is a bijection, and thus

$$\mathrm{End}_R(M^n) \cong M_n(\mathrm{End}_R(M))$$

are isomorphic rings. ■

2.2 Semisimple Rings

2.2.1 Ideals and Submodules

In this section, we are interested in proving structural theorems for semisimple rings, with the goal of obtaining useful descriptions of the various algebraic components of a ring. In particular, we will prove the Wedderburn and Artin theorems for semisimple rings, and then we will discuss their consequences and ramifications in detail.

Given a ring R, we say that it is semisimple if it is a semisimple R-module. That is, semisimple rings are precisely those that, when viewed as modules over themselves, can be written as a direct sum of simple R-submodules, which in this case are precisely their minimal left ideals. We will now prove a basic characterization for these rings.

Proposition 2.2.1. A ring R is semisimple if and only if every R-module is semisimple.

Proof. Let R be a ring, and note that since R is an R-module, the reverse of the statement follows immediately. Now assume that R is semisimple and take any R-module M. Consider the following R-homomorphism:

$$\varphi : \bigoplus_{m \in M} R \mapsto M;$$
$$\varphi((r_m)_{m \in M}) = \sum_{m \in M, r_m \neq 0} r_m \cdot m.$$

Note that for any $x \in M$, we can take the sequence $(\delta_{xm})_{m \in M}$, where $\delta_{xm} = 1$ if and only if $m = x$, and observe that $\varphi((\delta_{xm})_{m \in M}) = x$, so φ is surjective. Therefore, we have

$$\bigoplus_{m \in M} R/\ker(\varphi) \cong_R M,$$

and since the direct sum of semisimple modules is semisimple, we conclude that M is isomorphic to a quotient of a semisimple module. Therefore, by Corollary 2.1.3, it follows that M is semisimple. ■

Semisimplicity in rings also implies a stronger version of Brauer's Lemma.

Proposition 2.2.2. Let R be a semisimple ring. Then every left ideal of R is of the form Re, where $e^2 = e$ is an idempotent element. In particular, every bilateral ideal is a ring of the form Re, with the unit e belonging to the center of R.

Proof. Let L be a left ideal of R, and since R is semisimple, we know from Theorem 2.1.2 that there exists another left ideal J such that $R = L \oplus J$. Let $e \in L$ and $x \in J$ such that $e + x = 1$. For any $a \in L$, it follows that

$$a = ae + ax,$$

which implies that

$$ax = a - ae \in J \cap L,$$

but since $J \cap L = \{0\}$, it follows that $ax = 0$ and $a = ae$. This implies that $e^2 = e$, and also that $L = Re$, and a similar result holds for right ideals. Now assume that I is a bilateral ideal. We know that there are idempotent elements e, e' such that

$$I = Re = e'R,$$

so it follows that $ee = e'e$ and that $e'e = e'e'$. But since both are idempotents, it follows that $e = e'e = e'$, and hence $I = Re = eR$. Note that if $x \in I$, there exist $a_1, a_2 \in R$ such that

$$x = a_1 e = e a_2$$

so

$$xe = a_1 e = x \qquad \text{and} \qquad ex = e a_2 = x,$$

hence $xe = ex$, implying that e commutes with every element of I, i.e., I is a ring with unit e. If we take any element $a \in R$, we have

$$ae = (ae)e = eae \qquad \text{and} \qquad ea = e(ea) = eae,$$

so $ae = ea$, meaning that e is an element of the center of R, as required. ■

Now we show that the decomposition into simple submodules of a ring is always finite.

Proposition 2.2.3. If R is a semisimple ring, then it can be written as a finite direct sum of simple submodules. Moreover, every simple R-module is isomorphic to one of the direct summands of R.

Proof. Let $R = \bigoplus_i L_i$, where each L_i is a simple R-submodule. Note that, by definition, there exist unique $r_{i_1}, ..., r_{i_m} \in L_{i_j}$ such that

$$1 = r_{i_1} + \cdots + r_{i_m}.$$

Therefore, any element $x \in R$ can be uniquely written as

$$x = x r_{i_1} + \cdots + x r_{i_m},$$

and since each L_{i_j} is an R-submodule of R, we have $x r_{i_j} \in L_{i_j}$, which implies that $R = \bigoplus_{j=1}^{m} L_{i_j}$.

Now let M be a simple R-module. We know that any non-zero element m of M satisfies $M = Rm$, so the R-homomorphism $\varphi : R \mapsto M$ given by $\varphi(a) = am$ is non-zero. On the other hand, from Proposition 1.3.8, we have

$$\{0\} \neq \mathrm{Hom}_R(R, M) \cong \prod_{i=1}^{m} \mathrm{Hom}_R(L_i, M),$$

where the isomorphism is between abelian groups. But note that this implies that there exists some i such that $\mathrm{Hom}_R(L_i, M) \neq \{0\}$, so by Schur's Lemma, it follows that $M \cong_R L_i$. ■

2.2.2 Wedderburn's Theorem

An immediate consequence of the previous proposition is that there are finitely many isomorphism classes of simple modules in a semisimple ring. If we write a semisimple ring R as

$$R = \bigoplus_{i=1}^{l} L_i,$$

where each L_i is an R-simple submodule, we can then consider the submodules

$$R_i := \bigoplus_{L \cong_R L_i} L \cong_R L_i^{d_i}$$

which group all the d_i isomorphic summands to L_i in the direct sum of R, for some non-negative integer d_i. Thus, each R_i is a left ideal of R that is semisimple, and we can write

$$R = \bigoplus_{i=1}^{m} R_i \cong_R \bigoplus_{i=1}^{m} L_i^{d_i},$$

where $L_i \ncong_R L_j$ if $i \neq j$. We can also note that each R_i is a bilateral ideal of R. Indeed, if $i \neq j$, we take

$$R_i R_j = \bigoplus_{L \cong_R L_j} R_i L = \bigoplus_{L \cong_R L_j} \bigoplus_{L' \cong_R L_i} L'L.$$

We know that both L, L' are simple submodules, so if there exist $a \in L', b \in L$ such that $ab \neq 0$, the map sending any $a \in L'$ to $ab \in L$ would be a non-zero R-homomorphism, and therefore by Schur's Lemma, L and L' would be isomorphic, which contradicts the fact that $i \neq j$, hence $L'L = \{0\}$, implying that $R_i R_j = \{0\}$. From this, it follows that

$$R_i \subseteq R_i R = R_i \bigoplus_{j=1}^{m} R_j = \bigoplus_{j=1}^{m} R_i R_j = R_i R_i \subseteq R_i,$$

where the last inclusion follows from the fact that R_i is a left ideal of R, so $R_i R = R_i$, implying that R_i is a bilateral ideal. From this and by Proposition 2.2.2, it follows that each R_i is a ring with a unit given by some idempotent e_i belonging to the center of R, and since $R_i R_j = \{0\}$ for $i \neq j$, we have that $e_i e_j = 0$.

Since R is the direct sum of the R_i's, we have that the unit of R can be uniquely written as

$$1 = e_1 + \cdots + e_m.$$

Moreover, if we take elements $r, s \in R$, they can be uniquely decomposed as $r = r_1 + \cdots + r_m$ and $s = s_1 + \cdots + s_m$, with $r_i, s_i \in R_i$, so their multiplication is given by

$$rs = r_1 s_1 + \cdots + r_m s_m,$$

since the products of the form $r_i s_j$ with $i \neq j$ will be zero. Therefore, the canonical map that sends r to $(r_1, \ldots, r_m)$ — that is, which identifies the internal direct sum of submodules with the external direct sum — will also be a ring isomorphism, since the unit 1 will be mapped to $(e_1, \ldots, e_m)$, which is precisely the unit of the direct product of the R_i. This shows that the ring R is isomorphic to an external direct product of the rings R_i, and since each ring corresponds to a bilateral ideal of R, we say that

$$R = \prod_{i=1}^{m} R_i,$$

that is, we say that R is equal to the direct product of the R_i. This is one possible definition of internal direct products in rings, but since its construction is somewhat specific to the case of semisimple rings, we chose to introduce it only now.

This discussion leads us to the proof of Wedderburn's theorem.

Theorem 2.2.4 (Wedderburn's Theorem)**.** Let R be a semisimple ring written as

$$R = \bigoplus_{i=1}^{m} R_i,$$

where $R_i \cong_R L_i^{d_i}$, each L_i is a simple submodule, and $L_i \ncong_R L_j$ if $i \neq j$. Then each R_i is a simple ring isomorphic to $M_{d_i}(D_i)$, where D_i is a division ring, and we also have that

$$R \cong \prod_{i=1}^{m} M_{d_i}(D_i),$$

that is, R is isomorphic to a direct product of matrix rings over division rings. Conversely, every direct product of matrix rings over division rings is semisimple.

Proof. By Propositions 1.3.5 and 2.1.4, we have that

$$\mathrm{End}_R(R_i) \cong \mathrm{End}_R(L_i^{d_i}) \cong M_{d_i}(\mathrm{End}_R(L_i))$$

are isomorphic rings, so

$$\begin{aligned} R_i &\cong (R_i^{\mathrm{op}})^{\mathrm{op}} \\ &\cong (\mathrm{End}_R(R_i))^{\mathrm{op}} \\ &\cong (M_{d_i}(\mathrm{End}_R(L_i)))^{\mathrm{op}} \\ &\cong M_{d_i}(\mathrm{End}_R(L_i)^{\mathrm{op}}), \end{aligned}$$

where the second and last isomorphisms follow from Propositions 1.2.6 and 1.3.9, and the ring $D_i := \mathrm{End}_R(L_i)^{\mathrm{op}}$ is a division ring because L_i is a simple module. Therefore, the ring R_i is isomorphic to $M_{d_i}(D_i)$ for each i, and so it is a simple ring. Proposition 1.3.8 gives us that

$$\mathrm{End}_R(R) = \mathrm{End}_R(\bigoplus_{i=1}^{m} R_i) \cong \prod_{i,j} \mathrm{Hom}_R(R_i, R_j),$$

where the isomorphisms are of abelian groups, and by assumption, $L_i \ncong L_j$ if $i \neq j$, but note that the ideals L_i, L_j are simple R-modules and not isomorphic, so by Schur's Lemma, any R-homomorphism between L_i, L_j is identically zero, i.e., $\mathrm{Hom}_R(L_i, L_j) = \{0\}$, which also implies that $\mathrm{Hom}_R(R_i, R_j) = \{0\}$, so

$$\mathrm{End}_R(R) \cong \prod_{i=1}^{m} \mathrm{End}_R(R_i) \cong \prod_{i=1}^{m} M_{d_i}(\mathrm{End}_R(L_i)),$$

where the isomorphisms are now of rings. Now, if we proceed similarly to what was done for each R_i, we obtain

$$\begin{aligned} R &\cong (R^{\mathrm{op}})^{\mathrm{op}} \\ &\cong (\mathrm{End}_R(R))^{\mathrm{op}} \\ &\cong (\prod_{i=1}^{m} M_{d_i}(\mathrm{End}_R(L_i)))^{\mathrm{op}} \\ &\cong \prod_{i=1}^{m} (M_{d_i}(\mathrm{End}_R(L_i)))^{\mathrm{op}} \\ &\cong \prod_{i=1}^{m} M_{d_i}(\mathrm{End}_R(L_i)^{\mathrm{op}}) \\ &= \prod_{i=1}^{m} M_{d_i}(D_i). \end{aligned}$$

Thus, R is isomorphic to a direct product of matrix rings over division rings, where each term of the product is isomorphic to the corresponding R_i in the expression of R as a direct sum. Conversely, if R is a direct product of matrix rings over division rings, we know from Example 1.3.13 that each element of the product is semisimple, and since it is a direct product, each element will be a bilateral ideal, so we can identify R also as a direct sum of semisimple submodules, and from this, it follows that R is semisimple. ■

Wedderburn's theorem has several relevant implications for the study of rings and algebras, and we will dedicate the remainder of this chapter to discuss these consequences in detail. The most immediate and perhaps most important implication is precisely the guarantee that we can identify a semisimple ring R with a direct product of simple rings R_i, each of which is a bilateral ideal of R isomorphic to a matrix ring over a division ring. This decomposition of a semisimple ring as a product of simple rings R_i is commonly called the *Wedderburn decomposition* of the ring. We also observe that the theorem guarantees that it is not necessary to distinguish between left or right semisimplicity: matrix rings over division rings are semisimple both on the left (with minimal left ideals given by column matrices C_i described in Example 1.3.13) and on the right (with minimal right ideals given by row matrices defined analogously). That is, a ring is semisimple on the right if and only if it is semisimple on the left.

Now we show that the Wedderburn decomposition is unique.

Proposition 2.2.5 (Uniqueness of the Wedderburn Decomposition)**.** If R is a semisimple ring with Wedderburn decomposition given by

$$R = \prod_{i=1}^{m} R_i,$$

where each R_i is a simple ring, then this decomposition is unique up to permutation of the R_i, where the parameters m, d_i are uniquely determined by R.

Proof. Assume that

$$R = \prod_{i=1}^{m} R_i = \prod_{j=1}^{m'} R'_j,$$

and note that

$$R_i \subseteq R_i R = R_i R_i \subseteq R_i,$$

so $R_i R = R_i$ for any i. On the other hand, we also have

$$R_i R = \prod_{j=1}^{m'} R_i R'_j,$$

but since each R'_j is also a bilateral ideal of R, it follows that $R_i R'_j$ is a bilateral ideal of R_i, and since R_i is simple, this implies that $R_i R'_j = R_i$ or $R_i R'_j = \{0\}$. Since $R_i R = R_i$, it follows that there exists exactly one index j such that $R_i R'_j = R_i$, and if we now repeat the process for j, we conclude that

$$R_i = R_i R'_j = R'_j,$$

so for each i there exists a unique j' such that $R_i = R'_j$, and therefore the Wedderburn decomposition of R is unique up to permutation of the R_i's, and consequently the parameters R_i, d_i, m are uniquely determined by R. ■

The main consequence of the Wedderburn decomposition for this work is the structure of the idempotent elements of a semisimple ring R. If $e \in R$ is an idempotent element belonging to the center of R, we say it is a *central idempotent*, and if e_i, e_j are distinct idempotents of R such that $e_i e_j = 0$, we say they are *orthogonal*. We say that an idempotent $e \in R$ is *primitive* (or *minimal*) if it cannot be written as the sum of non-zero orthogonal idempotents distinct from R, and an idempotent is said to be *centrally primitive* if it is central and cannot be written as the sum of non-zero orthogonal central idempotents.

With this new terminology, we obtain the following corollary:

Corollary 2.2.6. Let R be a semisimple ring with Wedderburn decomposition given by

$$R = \prod_{i=1}^{m} R_i.$$

Then there exist unique orthogonal centrally primitive idempotents $e_1, \ldots, e_m$ such that

$$1 = e_1 + \cdots + e_m.$$

Moreover, for each e_i, there also exist unique orthogonal primitive idempotents $e_{i1}, \ldots, e_{id_i}$ such that

$$e_i = e_{i1} + \cdots + e_{id_i}.$$

Proof. The uniqueness of the idempotents follows from the uniqueness of the Wedderburn decomposition, and we have seen in previous discussions that the idempotents $e_1, \ldots, e_m$ such that $R_i = Re_i$ are indeed orthogonal and central, so it remains to show that they are centrally primitive. Fix i and assume that $e_i = e_i' + e_i''$, where e_i', e_i'' are non-zero orthogonal central idempotents, i.e., we can write the ring R_i as

$$R_i = Re_i = Re_i' \oplus Re_i'',$$

where the sum is direct because the idempotents are orthogonal, but note that since e_i' and e_i'' are central, the submodules Re_i' will be proper non-trivial bilateral ideals of R_i, contradicting the simplicity of R_i, so each e_i is indeed centrally primitive.

We recall that by definition each R_i is of the form

$$R_i = \bigoplus_{L \cong_R L_i} L \cong_R L_i^{d_i},$$

that is, R_i is a semisimple ring, and its d_i left ideals are all isomorphic to L_i. By Proposition 2.2.2, it follows that for each left ideal of R it is of the form Re_{ij} for some idempotent e_{ij}, and by a similar argument to the one made earlier, it follows that these idempotents are orthogonal and primitive. Therefore, since e_i is the unit in R_i, we have that

$$e_i = e_{i1} + \cdots + e_{id_i},$$

as we wanted. ∎

Thus, we can describe the case where R is a commutative semisimple ring.

Corollary 2.2.7. Every commutative semisimple ring R is a finite direct product of fields. Conversely, every finite direct product of fields is semisimple.

Proof. By Wedderburn's theorem, we can write

$$R \cong \prod_{i=1}^{m} M_{d_i}(D_i),$$

where D_i is a division ring. Since R is commutative, it follows that each $M_{d_i}(D_i)$ must also be commutative, but since it is a matrix ring, this is only possible if $d_i = 1$. Each $M_1(D_i)$ is naturally isomorphic to D_i as a ring, so each D_i is a commutative division ring, and therefore it is a field. Conversely, if R is a finite direct product of fields, note that each field can be seen as a matrix ring of size 1×1 with entries in the field, so it is semisimple, which implies that R is also semisimple. ∎

2.2.3 Simple Rings and the Artin-Wedderburn Theorem

Wedderburn's theorem gives us a classification of semisimple rings, and a natural question we might ask is: is it also possible to characterize simple rings? To address this, we first need to introduce the concept of *Artinian rings*. Given a ring R, we say that it is left Artinian if any descending chain of left ideals stabilizes, that is, for left ideals $\{L_j\}_j$ such that

$$L_1 \supseteq L_2 \supseteq \cdots \supseteq L_n \supseteq L_{n+1} \supseteq \ldots,$$

there exists an index k such that for all $j \geq k$, $L_j = L_k$. This is equivalent to saying that every non-empty family of left ideals of R has a minimal element. It is also interesting to note that if L is a non-zero left ideal of R, then it must contain some minimal left ideal; otherwise, we could construct a non-empty family of left ideals that does not have a minimal element.

At first glance, the notion of Artinian rings may seem somewhat arbitrary, but as is often the case in abstract algebra, this term merely formalizes a property of familiar structures such as finite-dimensional algebras. Every finite-dimensional algebra is an Artinian ring simply because every left ideal is a subspace, so any strict chain of ideals is a chain of subspaces with decreasing dimension, and since the dimension is finite, this chain certainly stabilizes. A counterexample is the ring of integers $\mathbb{Z}$, as the chain

$$2\mathbb{Z} \supsetneq 4\mathbb{Z} \supsetneq \cdots \supsetneq 2k\mathbb{Z} \supsetneq \ldots$$

does not stabilize, so it is not Artinian. With this, we are ready to characterize the simple Artinian rings.

Theorem 2.2.8 (Artin-Wedderburn Theorem)**.** Let R be a ring. The following statements are equivalent:

(1) The ring R is simple and Artinian;

(2) There exists a simple and faithful R-module M;

(3) R is semisimple and has a unique isomorphism class of simple R-modules;

(4) R is isomorphic to $M_n(D)$, for some natural number n and a division ring D.

Proof.

(1) $\Rightarrow$ (2) We know that any ring R has a maximal left ideal L, and therefore the R-module R/L is simple. Note that $\mathrm{Ann}_R(R/L)$ is a bilateral ideal of R, but since R is simple and has a unit, it follows that $\mathrm{Ann}_R(R/L) = \{0\}$, so $M = R/L$ is simple and faithful.

(2) $\Rightarrow$ (3) Let M be a simple and faithful R-module, and consider the family of left ideals of R given by

$$\mathcal{F} := \{\ker(\varphi) \mid \varphi \in \mathrm{Hom}_R(R, M^n), n \in \mathbb{N}\}.$$

Since R is Artinian, it follows that there exists a natural number n and $\varphi \in \mathrm{Hom}_R(R, M^n)$ such that $\ker(\varphi)$ is minimal. We will show that φ is injective. In fact, if it were not, take a non-zero element $r \in R$ such that $\varphi(r) = 0$, and note that since M is faithful, there exists $m \in M$ such that $rm \neq 0$, i.e., $r \notin \mathrm{Ann}_R(m)$. We can then consider the map given by

$$\begin{aligned} \psi : R &\mapsto M^n \oplus M, \\ \psi(r) &= (\varphi(m), rm). \end{aligned}$$

Note that ψ is an R-homomorphism such that $\ker(\psi) = \ker(\varphi) \cap \mathrm{Ann}_R(m) \subsetneq \ker(\varphi)$, which contradicts the minimality of φ. Therefore, we conclude that φ is injective, and hence $R \cong_R \mathrm{Im}(\varphi)$. Since M is simple, it follows that M^n is semisimple, which in turn implies that $\mathrm{Im}(\varphi)$ is also semisimple, allowing us to conclude that R is semisimple and has a unique isomorphism class of simple R-modules.

(3) $\Rightarrow$ (4) If R is semisimple and has a unique isomorphism class of simple R-modules, it follows that the Wedderburn decomposition of R has a single component, and therefore Wedderburn's theorem guarantees that $R \cong M_n(D)$, for some natural number n and a division ring D.

(4) $\Rightarrow$ (1) We have already seen that $M_n(D)$ is a simple ring, and it is also an Artinian ring because $M_n(D)$ is a finite-dimensional vector space over D. ■

The previous theorem shows that a simple Artinian ring is always isomorphic to a matrix ring over a division ring, and with this, we conclude the main results on the semisimplicity of rings.

2.3 The Jacobson Radical

Now we will have a brief discussion on an important tool for the study of semisimple rings and algebras, known as the Jacobson radical of a ring. We define the *Jacobson Radical* $J(R)$ of a ring R as the intersection of all left maximal ideals, that is,

$$J(R) := \bigcap\{L \subseteq R | L \text{ is a left maximal ideal}\}.$$

This object will be useful for us because it is possible to characterize the semisimplicity of a given ring in terms of its radical, and we will see that in some cases it is significantly easier to demonstrate that an algebra is semisimple by studying its radical.

First, observe that if R is a ring and $y \in J(R)$ is an element of its radical, then $1+y$ must be left-invertible, that is, there exists $x \in R$ such that $x(1+y) = 1$. In fact, if $1+y$ were not invertible, then the left ideal $R(1+y)$ would be non-trivial and proper in R, so it would be contained in some left maximal ideal, but since $y \in J(R)$ and $J(R)$ is contained in all left maximal ideals, it follows that $1-y+y = 1$ would belong to some left maximal ideal of R, which is a contradiction. This also implies that any element of the form $1-y$ with $y \in J(R)$ is left-invertible in R.

The definition of the radical also allows us to conclude the following facts:

Proposition 2.3.1. If R is a ring, then:

(1) $J(R)$ is a two-sided ideal;

(2) If M is a simple R-module, then $J(R)M = \{0\}$. In particular, we have that

$$J(R) = \bigcap\{\text{Ann}_R(M) | M \text{ is a simple } R\text{-module}\};$$

(3) The radical of the ring $R/J(R)$ is $\{0\}$;

(4) If R is Artinian, its radical is a finite intersection of left maximal ideals.

Proof.

(1) Let $x \in J(R)$ and $r \in R$, and let L be any left maximal ideal of R. Note that by definition, $x \in L$. Define the map

$$\varphi : R/L \mapsto R/L,$$
$$\varphi(a+L) = ar + L,$$

and note that φ is an R-endomorphism, since for $m \in R$, we have $\varphi(ma+L) = mar+L = m(ar+L)$, and $\varphi((a+b)+L) = (a+b)r+L = (ar+L)+(br+L)$. We know that any R-homomorphism maps the zero of the domain to the zero of the codomain, so since $x \in L$, we have $x+L = 0$, thus $\varphi(x+L) = xr+L = 0$, and hence $xr \in L$. Therefore, xr belongs to any left maximal ideal of R, and so $xr \in J(R)$, proving that the radical is a two-sided ideal.

(2) If M is a simple R-module, we know from Theorem 1.3.11 that there exists a left maximal ideal L of R such that $M \cong_R R/L$. From the previous part, we know that $J(R)$ is a two-sided ideal, so if $x \in J(R)$, then $xr \in J(R)$ for any $r \in R$, and by definition, $J(R) \subseteq L$, so $xr \in L$, which implies that $xr + L = 0$, hence

$$J(R)M = \{0\},$$

implying that $J(R) \subseteq \text{Ann}_R(M)$. On the other hand, if x is an element of R that belongs to the intersection of all annihilators of simple R-modules, it follows that $x(1+L) = 0$ in the quotient R/L for any left maximal ideal L of R, so $x \in L$, and hence x is an element of $J(R)$. Thus, we conclude that

$$J(R) = \bigcap \{\text{Ann}_R(M) | M \text{ is a simple } R\text{-module}\},$$

as desired.

(3) The Correspondence Theorem for rings tells us that there is a bijection between the left ideals of R containing $J(R)$ and the left ideals of $R/J(R)$. Therefore, if $x + J(R)$ belongs to the radical of $R/J(R)$, it follows that $x + J(R)$ belongs to all left maximal ideals of $R/J(R)$, and consequently x belongs to all left maximal ideals of R containing $J(R)$ – which by definition are all the left maximal ideals of R –, so $x \in J(R)$. Hence, $x + J(R) = 0$, and thus the radical of $R/J(R)$ is $\{0\}$.

(4) Let R be Artinian, and consider the set

$$\mathcal{F} = \{\bigcap_{i=1}^{n} L_i | L_i \text{ is a left maximal ideal of } R, n \in \mathbb{N}\}.$$

Since R is Artinian, $\mathcal{F}$ has a minimal element $L = \bigcap_{i=1}^{n} L_i$ for some natural n. If $J(R) \neq L$, then there exists a left maximal ideal $I_j \nsubseteq L$, but then $L_j \in \mathcal{F}$ and $L_j \cap L \subsetneq L$, contradicting the minimality of L. Thus, $J(R)$ is a finite intersection of left maximal ideals.

■

The Jacobson radical is also strongly connected to the concept of nilpotency. We say that an ideal L – unilateral or bilateral – of a ring is *nilpotent* if there exists a natural number r such that $L^{(r)} = \{0\}$, in other words, any product of r elements of L is zero. This allows us to prove an important result.

Proposition 2.3.2. If R is a ring, then the following hold:

(1) Any nilpotent two-sided ideal of R is contained in $J(R)$;

(2) If R is Artinian, then $J(R)$ is nilpotent.

Proof.

(1) Let N be a nilpotent ideal of R, and let L be any left maximal ideal. Then $M = R/L$ is a simple R-module, so the R-submodule NM of M is either $\{0\}$ or M. If $NM = \{0\}$, then since $1 \in R$, all elements of the form $a + L$, with $a \in N$, are zero, so $N \subseteq L$. If $NM = M$, we repeat the process until we reach $N^{(r)}M = M$, where r is the natural number such that $N^{(r)} = \{0\}$, and in this case, if $N \nsubseteq L$, we would have $M = \{0\}$, a contradiction, so N is contained in every left maximal ideal of R, and therefore contained in its radical.

(2) Let J be the Jacobson radical of R, and consider the following descending chain of ideals

$$J \supseteq J^{(2)} \supseteq ...$$

Since R is Artinian, there exists $r \in \mathbb{N}$ such that $J^{(r)} = J^{(r+i)}$, for any $i \in \mathbb{N}$. We claim that $J^{(r)} = \{0\}$. In fact, if it were not, we would have $J^{(r)}J^{(r)} = J^{(r)} \neq \{0\}$, so there are non-zero elements $a, b \in J^{(r)}$ such that $ab \neq 0$, and hence the left ideal $J^{(r)}b \neq \{0\}$. Thus, it follows that $J^{(r)}J^{(r)}b = J^{(r)}b \neq 0$, so the set of left ideals L of $J^{(r)}$ such that $J^{(r)}L \neq \{0\}$ is non-empty, and since R is Artinian, this implies that there exists a minimal left ideal L such that $J^{(r)}L \neq \{0\}$. Certainly, such an ideal is principal, since there exists a non-zero element $x \in L$ such that $J^{(r)}x \neq \{0\}$, so $L = J^{(r)}x$. In particular, we can then find an element $y \in J^{(r)}$ such that $x = yx$, so $(1-y)x = 0$, but since $y \in J$, it follows that $1 - y$ is left-invertible, and hence $x = 0$. This implies that $L = J^{(r)}x = \{0\}$, contradicting the fact that L was minimal, and therefore we must have $J^{(r)} = \{0\}$.

■

Note that the previous result shows us that the Jacobson radical is precisely the largest nilpotent two-sided ideal of an Artinian ring, and we will now see that an Artinian ring is semisimple if and only if its radical is trivial.

Theorem 2.3.3. Let R be an Artinian ring. Then R is semisimple if and only if $J(R) = \{0\}$.

Proof.

($\Rightarrow$) If R is semisimple, we know that there are unique orthogonal idempotent elements $e_1, ..., e_m$ such that

$$1 = e_1 + ... + e_m,$$

where $R_i = Re_i$ correspond to the simple rings in the Wedderburn decomposition of R. Therefore, if $a \in J(R)$, we have that

$$a = a(e_1 + ... + e_m) = ae_1 + ... + ae_m,$$

but we also know that each e_i can be uniquely decomposed as a sum of orthogonal primitive idempotents $e_{i1}, ..., e_{id_i}$, where each e_{ij} is an idempotent of a minimal left ideal of R isomorphic to L_i. Therefore, we obtain that

$$ae_i = ae_{i1} + ... + ae_{id_i},$$

and from Proposition 2.3.1, we know that $J(R)L_i = \{0\}$ since L_i is a simple submodule, so $ae_{ij} = 0$ for any i and j, implying that $a = 0$ and thus $J(R) = \{0\}$.

($\Leftarrow$) From the previous proposition, we know that the radical of R is $\bigcap_{i=1}^{n} L_i$ for some natural n, where each L_i is a left maximal ideal. Therefore, consider the following map

$$\varphi : R \mapsto \bigoplus_{i=1}^{n} R/L_i,$$
$$\varphi(r) = (r + R/L_1, ..., r + R/L_n).$$

Note that this map is clearly a ring homomorphism, and if $\varphi(r) = 0$, then $r \in L_i$ for all i, so $r \in J(R) = \{0\}$, implying that the map is injective. It is easy to see that it is also surjective, and therefore φ is an isomorphism of rings between R and a finite direct sum of R-modules, each of which is simple since each L_i is a left maximal ideal, so R is semisimple. ■

Example 2.3.4. We know that the ring of matrices $M_n(\mathbb{R})$ is simple and Artinian, and since its radical J is a proper two-sided ideal, it follows that $J = \{0\}$. On the other hand, if we consider the subring $\mathrm{UT}_n(\mathbb{R})$ of upper triangular matrices, the subset of matrices with zeros on the diagonal will be a nilpotent two-sided ideal, so $J(\mathrm{UT}_n(\mathbb{R})) \neq \{0\}$, and thus $\mathrm{UT}_n(\mathbb{R})$ is not semisimple.

2.4 Semisimple Algebras

Given an algebra $\mathcal{A}$ over a field $\mathbb{F}$, we say that it is *semisimple* if it is semisimple as a ring, that is, if $\mathcal{A}$ can be written as a direct sum of simple $\mathcal{A}$-submodules. Similarly, we say that an algebra is *simple* if it is simple as a ring, i.e., if its only non-trivial bilateral ideal is the zero ideal.

We consider algebras of finite dimension over algebraically closed fields, such as $\mathbb{C}$, for example, to conclude a stronger version of Schur's Lemma.

Lemma 2.4.1. Let $\mathcal{A}$ be a finite-dimensional algebra over an algebraically closed field $\mathbb{F}$, and let $\mathcal{L}$ be a simple $\mathcal{A}$-submodule. Then $\mathrm{End}_{\mathcal{A}}(\mathcal{L})$ and $\mathbb{F}$ are isomorphic fields.

Proof. First, note that by definition, every $\mathcal{A}$-submodule is a vector space, that is, it has finite dimension. Moreover, given an element $f \in \mathrm{End}_{\mathcal{A}}(\mathcal{L})$, note that f is also an $\mathbb{F}$-endomorphism because for any $\alpha \in \mathbb{F}$, and for any $A \in \mathcal{L}$, we have

$$f(\alpha A) = f(\alpha EA) = f((\alpha E)A) = \alpha E f(A) = \alpha f(A),$$

hence $\mathrm{End}_{\mathcal{A}}(\mathcal{L}) \subseteq \mathrm{End}_{\mathbb{F}}(\mathcal{L})$. Since $\mathbb{F}$ is algebraically closed, there exists an eigenvalue $\lambda \in \mathbb{F}$ of f with a non-zero eigenvector $X \in \mathcal{L}$ such that

$$f(X) = \lambda X.$$

Therefore, the map $f - \lambda E$ is an $\mathbb{F}$-endomorphism with a non-trivial kernel, which belongs to $\mathrm{End}_{\mathcal{A}}(\mathcal{L})$. However, since $\mathcal{L}$ is a simple $\mathcal{A}$-module, Schur's Lemma implies that $\mathrm{End}_{\mathcal{A}}(\mathcal{L})$ is a division ring, so $f - \lambda E$ must be identically zero, which implies $f = \lambda E$. Now consider the following ring homomorphism:

$$\begin{aligned} \varphi : \mathbb{F} &\mapsto \mathrm{End}_{\mathcal{A}}(\mathcal{L}), \\ \varphi(\lambda) &= \lambda E. \end{aligned}$$

Note that φ is clearly injective, since $\lambda E = 0$ implies that $\lambda = 0$. The previous observations about $\mathrm{End}_{\mathcal{A}}(\mathcal{L})$ show that φ is surjective, thus the result follows. ■

Similarly to what was done for the case of rings, we can write a semisimple algebra $\mathcal{A}$ as

$$\mathcal{A} = \bigoplus_{i=1}^{l} \mathcal{L}_i,$$

where each $\mathcal{L}_i$ is a simple $\mathcal{A}$-submodule, and then we can consider the submodules

$$\mathcal{A}_i := \bigoplus_{\mathcal{L} \cong_{\mathcal{A}} \mathcal{L}_i} \mathcal{L} \cong_{\mathcal{A}} \mathcal{L}_i^{d_i}$$

which group all the d_i summands isomorphic to $\mathcal{L}_i$ in the direct sum of $\mathcal{A}$, for some non-negative integer d_i. Thus, we can write

$$\mathcal{A} = \bigoplus_{i=1}^{m} \mathcal{A}_i \cong_{\mathcal{A}} \bigoplus_{i=1}^{m} \mathcal{L}_i^{d_i},$$

where $\mathcal{L}_i \ncong_{\mathcal{A}} \mathcal{L}_j$ if $i \neq j$. From this, it follows that each $\mathcal{A}_i$ is a bilateral ideal of $\mathcal{A}$ of the form $\mathcal{A}_i = \mathcal{A}E_i$, where each E_i is a central idempotent of $\mathcal{A}$ that is also the identity of the algebra $\mathcal{A}_i$. This immediately implies that if $i \neq j$, then $\mathcal{A}_i\mathcal{A}_j = \{0\}$. Note that Wedderburn's Theorem guarantees that each $\mathcal{A}_i$ is isomorphic as a ring to $M_{d_i}(\mathrm{End}_R(\mathcal{L}_i)^{\mathrm{op}})$, and we leave it to the reader to verify that this isomorphism is also one of algebras. Since each $\mathcal{L}_i$ is a simple submodule of $\mathcal{A}$, Lemma 2.4.1 guarantees that $\mathrm{End}_R(\mathcal{L}_i)$ is a field isomorphic to $\mathbb{F}$, and therefore

$$\mathcal{A}_i \cong M_{d_i}(\mathbb{F})$$

are $\mathbb{F}$-algebras isomorphic to each other, and thus we can identify $\mathcal{A}$ with the direct product of the algebras $\mathcal{A}_i$ given by

$$\mathcal{A} = \prod_{i=1}^{m} \mathcal{A}_i.$$

These observations allow us to conclude analogous versions of the results discussed in the previous sections for algebras.

Theorem 2.4.2 (Wedderburn's Theorems for Algebras)**.** Let $\mathcal{A}$ be a semisimple finite-dimensional algebra over an algebraically closed field $\mathbb{F}$, written as

$$\mathcal{A} = \bigoplus_{i=1}^{m} \mathcal{A}_i,$$

where $\mathcal{A}_i \cong_{\mathcal{A}} \mathcal{L}_i^{d_i}$, each $\mathcal{L}_i$ is a simple submodule, and $\mathcal{L}_i \ncong_{\mathcal{A}} \mathcal{L}_j$ if $i \neq j$. Then the following hold:

(1) Each $\mathcal{A}_i$ is isomorphic as an $\mathbb{F}$-algebra to the full matrix algebra $M_{d_i}(\mathbb{F})$, and therefore

$$\mathcal{A} \cong \prod_{i=1}^{m} M_{d_i}(\mathbb{F})$$

is an isomorphism of $\mathbb{F}$-algebras. Conversely, any finite direct product of full matrix algebras is semisimple;

(2) The Wedderburn decomposition of the algebra $\mathcal{A}$ is unique up to permutation of the $\mathcal{A}_i$, and the parameters m, d_i are uniquely determined by $\mathcal{A}$;

(3) There exist unique centrally primitive orthogonal idempotents $E_1, ..., E_m$ such that

$$E = E_1 + ... + E_m,$$

where E is the identity of the algebra $\mathcal{A}$. Moreover, for each E_i, there also exist unique centrally primitive orthogonal idempotents $E_{i1}, ..., E_{id_i}$ such that

$$E_i = E_{i1} + ... + E_{id_i};$$

(4) If $\mathcal{A}$ is also commutative, then

$$\mathcal{A} \cong \prod_{i=1}^{m} \mathbb{F},$$

and in particular, the centrally primitive orthogonal idempotents form a basis for $\mathcal{A}$ as a $\mathbb{F}$-vector space.

■

Statement (1) is nothing more than the analogous version of Theorem 2.2.4, (2) is the analogous version of Proposition 2.2.5, and (3), (4) are the analogues of Corollaries 2.2.6 and 2.2.7, respectively, noting that in the case of commutative semisimple algebras, we will have that $\mathcal{A}$ will be isomorphic to $\prod_{i=1}^{m} \mathbb{F}$, which is a vector space of dimension m, and also noting that the set of idempotents $E_1, \ldots, E_m$ is linearly independent, it follows that they form a basis for $\mathcal{A}$.

It is interesting to note that the dimension of $\mathcal{A}$ will be precisely given by the sum $\sum_{i=1}^{m} d_i^2$, i.e., the dimension of $\mathcal{A}$ is always a sum of squares, and if we fix an integer n, then the possible partitions of n as a sum of squares will determine all possible semisimple subalgebras of $M_n(\mathbb{C})$, for example. It is also worth noting that the center $Z(\mathcal{A})$ of $\mathcal{A}$ can be written as

$$Z(\mathcal{A}) = \prod_{i=1}^{m} Z(\mathcal{A}_i) \cong \prod_{i=1}^{m} Z(M_{d_i}(\mathbb{F})) \cong \prod_{i=1}^{m} \mathbb{F},$$

since if $i \neq j$, then $\mathcal{A}_i\mathcal{A}_j = \{0\}$, and the center of $M_{d_i}(\mathbb{F})$ consists of multiples of the identity matrix. Therefore, in this case, the center of $\mathcal{A}$ will be semisimple, and its dimension will be precisely the number of isomorphism classes of simple $\mathcal{A}$-submodules.

We can now answer the motivating question posed at the beginning of this chapter: when is it possible to block-diagonalize a set of matrices simultaneously? The techniques developed so far allow us to conclude that a matrix algebra can be block-diagonalized if and only if it is semisimple. To this end, let $\mathcal{A} \subseteq M_n(\mathbb{C})$ – here we could consider any algebraically closed field – be a semisimple algebra with identity E written as

$$\mathcal{A} \cong_{\mathcal{A}} \bigoplus_{i=1}^{m} \mathcal{L}_i^{d_i},$$

where each $\mathcal{L}_i$ is a simple $\mathcal{A}$-module, and $\mathcal{L}_i \ncong \mathcal{L}_j$ if $i \neq j$. We saw at the end of the previous chapter that the simple submodules $\mathcal{L}_i$ of $\mathcal{A}$ are isomorphic as $\mathcal{A}$-modules to $\mathbb{C}^{d_i}$, i.e., the dimension of $\mathcal{L}_i$ as a $\mathbb{C}$-vector space is precisely d_i. Now, note that the identity matrix I can be written as

$$I = E + (I - E),$$

where

$$(I - E)^2 = I - E \quad \text{and} \quad E(I - E) = 0,$$

i.e., E and $I - E$ are two orthogonal projections that sum to I, implying that we can decompose the space $\mathbb{C}^n$ as the direct sum of their images, that is,

$$\mathbb{C}^n = E\mathbb{C}^n \oplus (I - E)\mathbb{C}^n.$$

Since E is the identity of $\mathcal{A}$, it follows that $E\mathbb{C}^n$ is an $\mathcal{A}$-module, and that $\mathcal{A}(I-E)\mathbb{C}^n = \{0\}$, i.e., both summands are $\mathcal{A}$-invariant, which means that if we consider a basis for $\mathbb{C}^n$ formed by the union of the bases for the summands, we obtain a basis where all matrices of $\mathcal{A}$ are block-diagonal, with one block corresponding to $E\mathbb{C}^n$, and one block with all entries equal to zero corresponding to $(I-E)\mathbb{C}^n$. Since $E\mathbb{C}^n$ is an $\mathcal{A}$-module, it follows from Proposition 2.2.1 that it is semisimple, i.e., we can decompose $E\mathbb{C}^n$ as a direct sum of simple $\mathcal{A}$-submodules, and from Proposition 2.2.3, it follows that each summand will be isomorphic to one of the $\mathcal{L}_i$ as an $\mathcal{A}$-module, and consequently as a $\mathbb{C}$-vector space. Then there exist non-negative integers k_i for every $i \in \{1, \ldots, m\}$ such that

$$E\mathbb{C}^n \cong_{\mathcal{A}} \bigoplus_{i=1}^{m} \mathcal{L}_i^{k_i},$$

and since each $\mathcal{L}_i$ is a $\mathbb{C}$-subspace of dimension d_i, if we consider a basis for $E\mathbb{C}^n$ composed of the union of bases for each of the subspaces isomorphic to some $\mathcal{L}_i$, we obtain that the block corresponding to $E\mathbb{C}^n$ will be in a block-diagonal form, with k_i blocks of size $d_i \times d_i$ for each i. This allows us to conclude that we can always find an invertible matrix P such that $P^{-1}\mathcal{A}P$ is a block-diagonal matrix algebra, i.e., the semisimple matrix algebras are indeed those for which it is possible to find a simultaneous block-diagonalization of all their elements.

2.5 Representations of groups and algebras

2.5.1 Initial Definitions

To conclude this chapter, we will present a brief discussion on representations of finite groups and algebras. The theory of representations is extremely deep, with relevant applications in areas such as harmonic analysis, quantum physics, Lie theory, and many others. For this reason, we believe it is interesting to present some basic results on the subject, and especially to discuss the connection this theory has with the notion of semisimplicity.

We begin with groups. If T is a homomorphism from a finite group G to the group $GL(V)$ of invertible linear transformations on V, where V is a vector space of dimension n over a field $\mathbb{F}$, we say that (T, V) is a linear representation of *degree* n over $\mathbb{F}$, and in this case, we also write $T_g = T(g)$ when convenient. Since V has finite dimension, we can fix a basis $v_1, \ldots, v_n$ for V and map each linear transformation to its corresponding matrix. The composition of this isomorphism with the representation will then give us a representation of G in $\mathrm{GL}(n, \mathbb{F})$, called the *matrix representation*.

The following example is fundamental for understanding the relationship between permutation and linear representations:

Example 2.5.1. Let S_n be the symmetric group on n elements, and let V be a vector space over $\mathbb{F}$ of dimension n. Fix a basis $v_1, \ldots, v_n$ for V, and define the endomorphism P_σ for $\sigma \in G$ as

$$P_\sigma v_i = v_{\sigma(i)},$$

that is, P_σ permutes the vectors of the basis according to σ. The map $\sigma \mapsto P_\sigma$ is then a faithful representation of S_n of degree n, and in particular, if we take $V = \mathbb{F}^n$, we obtain that this representation gives us an isomorphism between the elements of S_n and the group of permutation matrices in $\mathrm{GL}(n, \mathbb{F})$.

If G is a group with n elements and V has dimension n, we can then define the map

$$T : G \mapsto \mathrm{GL}(V),$$
$$T(g) = P_{g_L},$$

that is, we map the element g to the permutation matrix associated with the permutation g_L in S_n, where $g_L(h) = gh$ for any $h \in G$. This map is the composition of the monomorphism $g \mapsto g_L$ given by Cayley's Theorem with the monomorphism $g_L \mapsto P_{g_L}$ from the previous example. Thus, it gives us a faithful representation of G in $\mathrm{GL}(V)$, called the *regular representation* of the group, and naturally, we obtain the

regular matrix representation when we take $V = \mathbb{F}^n$. We say that two representations $(T, V), (T', V)$ of G are *equivalent* if there exists $S \in \mathrm{GL}(V)$ such that

$$T'_g = ST_gS^{-1},$$

for any $g \in G$. In terms of matrix representations, this means that the matrices of T' are obtained from a change of basis of the matrices of T.

2.5.2 Irreducibility and Maschke's Theorem

If (T, V) is a representation of G, we say that a subspace W of V is a *G-subspace* if W is invariant under all the matrices of the representation, that is, for any $g \in G$, it follows that $T_gW \subseteq W$. In this case, we can decompose the space V as a direct sum of W and its complement, and then $(T|_W, W)$ is a representation of G. We say that a representation T is *irreducible* if the only proper G-subspace of V is $\{0\}$, and otherwise, we say that T is *reducible*. A *completely reducible* representation is one in which every G-subspace W has a G-subspace complement W' such that $V = W \oplus W'$. With this, we can prove two important results.

Proposition 2.5.2. Let G be a finite group and (T, V) a completely reducible representation of G of degree n. Then the following hold:

(1) Every G-subspace induces a completely reducible representation;

(2) The space V can be decomposed as a direct sum of irreducible G-subspaces.

Proof.

(1) Let W be a G-subspace, and let $T|_W$ be the representation induced by W. Note that if N is a G-subspace of W, then it is also a G-subspace of V, and since T is completely reducible, it follows that we can find a G-subspace M of V such that

$$V = N \oplus M,$$

so

$$W = N \oplus (W \cap M),$$

but note that $W \cap M$ is also a G-subspace, so $T|_W$ is completely reducible.

(2) Let W be a proper and non-trivial G-subspace of V – if none exists, then V is irreducible, and there is nothing to prove. We can then find another G-subspace W' such that

$$V = W \oplus W',$$

where $\dim(W), \dim(W') < \dim(V) = n$. We can then use induction on n and the previous item to find direct sum decompositions into irreducible G-subspaces for W and W', and hence we obtain the desired result.

■

Now we can prove one of the most important basic results in group representation theory.

Theorem 2.5.3 (Maschke's Theorem)**.** If G is a finite group and (T, V) is a representation of G over $\mathbb{F}$ such that the characteristic $\mathrm{char}(\mathbb{F})$ of the field does not divide $|G|$, then T is completely reducible.

Proof. Let W be a G-subspace of V, and let W' be a complement of W in V – initially, W' may not be a G-subspace. We then obtain a decomposition

$$V = W \oplus W',$$

and then we can find a projector E in W such that $EV = W$, $E^2 = E$. Now define the operator

$$\mathcal{R} : \mathrm{End}_{\mathbb{F}}(V) \mapsto \mathrm{End}_{\mathbb{F}}(V),$$
$$\mathcal{R}(A) = \frac{1}{|G|} \sum_{g \in G} T_g A T_g^{-1},$$

known as the *Reynolds operator*. In the next chapters, we will see that this operator has several interesting properties and can be used to demonstrate some results in combinatorial optimization. First, we note that if $h \in G$ and $A \in \mathrm{End}_{\mathbb{F}}(V)$, then

$$\begin{aligned} T_h \mathcal{R}(A) T_h^{-1} &= \frac{1}{|G|} \sum_{g \in G} (T_h T_g) A (T_g^{-1} T_h^{-1}) \\ &= \frac{1}{|G|} \sum_{g \in G} T_{hg} A T_{hg}^{-1} \\ &= \mathcal{R}(A), \end{aligned}$$

where the last equality follows from the fact that the map $g \mapsto hg$ is a bijection in G, i.e., $G = hG$. Therefore, we conclude that the image of $\mathcal{R}$ is contained in the center of G in $\mathrm{End}_{\mathbb{F}}(V)$. Now note that if $w \in W$, then $T_g^{-1} w \in W$, so

$$T_g E T_g^{-1} w = T_g T_g^{-1} w = w,$$

which implies that $\mathcal{R}(E)W = W$. On the other hand, if $v \in W'$, then $E T_g^{-1} v \in W$, so $T_g E T_g^{-1} v \in W$, and thus $\mathcal{R}(E)V = W$, and therefore $\mathcal{R}(E)$ is indeed a projection onto W. We can then decompose V as

$$V = \mathcal{R}(E)V \oplus (I - \mathcal{R}(E))V,$$

and observe that since $\mathcal{R}(E)$ commutes with all the T_g, we have

$$T_g(I - \mathcal{R}(E)) = (I - \mathcal{R}(E))T_g,$$

so $(I - \mathcal{R}(E))V$ is also a G-subspace, and therefore T is completely reducible, as desired. ■

2.5.3 Group Algebra and Semisimplicity

Maschke's theorem, together with item (2) of the previous proposition, allows us to conclude that if T is a representation of a finite group G in $\mathrm{GL}(n, \mathbb{C})$, for example, and if $V = \mathbb{C}^n$ is decomposed as

$$V = W_1 \oplus \ldots \oplus W_k,$$

with each $T|_{W_i}$ irreducible, then the matrix T_g is in the form

$$T_g = \begin{pmatrix} T_g|_{W_1} & 0 & \ldots & 0 \\ 0 & T_g|_{W_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & T_g|_{W_k} \end{pmatrix},$$

where each $T_g|_{W_i}$ is a block of size $\dim(W_i) \times \dim(W_i)$, i.e., we can block-diagonalize all the matrices of the representation T. This result is extremely similar to what we obtain as a consequence of Wedderburn's Theorem for semisimple algebras, and this is no coincidence. There is a very natural way to unify the language used to describe representations with the concepts we have seen so far about simple submodules and semisimplicity, but for this, we first need to associate an algebra to the group G.

Definition 2.5.4 (Group Algebra)**.** If G is a finite group and $\mathbb{F}$ is a field, we define the *group algebra* as the set $\mathbb{F}G = \{\sum_{g\in G} \alpha_g g | \alpha_g \in \mathbb{F}\}$ of formal sums over G with coefficients in $\mathbb{F}$, with operations given by

$$\begin{aligned}
\beta \cdot (\sum_{g\in G} \alpha_g g) &= \sum_{g\in G} (\beta\alpha_g) g, \\
(\sum_{g\in G} \alpha_g g) + (\sum_{g\in G} \beta_g g) &= \sum_{g\in G} (\alpha_g + \beta_g) g, \\
(\sum_{g\in G} \alpha_g g) \cdot (\sum_{h\in G} \beta_h h) &= \sum_{g\in G} (\sum_{h\in G} \alpha_{gh^{-1}} \beta_h) g
\end{aligned}$$

for any $\beta \in \mathbb{F}$ and $\sum_{g\in G} \alpha_g g, \sum_{h\in G} \beta_h h \in \mathbb{F}G$.

The scalar multiplication and addition operations make $\mathbb{F}G$ a vector space of dimension $|G|$, where the set of elements $g = 1 \cdot g$ forms a basis that can naturally be identified with G. The multiplication operation may seem complicated, but we can derive it as follows: first, we write the product of two elements in the usual form,

$$(\sum_{g\in G} \alpha_g g) \cdot (\sum_{h\in G} \beta_h h) = \sum_{g,h\in G} \alpha_g \beta_h gh,$$

and then we perform the change of variables $t = gh$, i.e., $g = th^{-1}$, and we get

$$\sum_{g,h\in G} \alpha_g \beta_h gh = \sum_{g\in G} (\sum_{h\in H} \alpha_{th^{-1}} \beta_h) t,$$

but note that since $t = gh$ and right multiplication by h is a bijection from G to G, it follows that

$$\sum_{g\in G} (\sum_{h\in H} \alpha_{th^{-1}} \beta_h) t = \sum_{t\in G} (\sum_{h\in H} \alpha_{th^{-1}} \beta_h) t,$$

and then it is enough to rename the variable t to obtain the expression from the definition. Another interesting way to view the group algebra is through functions from G to $\mathbb{F}$, that is, we can consider the set $\{\varphi : G \mapsto \mathbb{F}\}$ and then take the functions δ_g defined by $\delta_g(h) = 1$ if $h = g$ and 0 otherwise, so any $\varphi : G \mapsto \mathbb{F}$ can be written as

$$\varphi = \sum_{g\in G} \varphi(g) \delta_g,$$

and then we just need to identify δ_g with g to obtain $\mathbb{F}G$. In this context, multiplication becomes a convolution, because

$$(\varphi \cdot \psi)(g) = \sum_{h\in H} \varphi(gh^{-1}) \psi(h).$$

With these definitions, it is clear that $\mathbb{F}G$ is an $\mathbb{F}$-algebra with unit $e = 1 \cdot e$, where e is the identity element of G, and we will now see that much of the structural information of the group G is encoded in the algebraic structure of $\mathbb{F}G$. We will now define the notion of a representation for any algebra. If $\mathcal{A}$ is a finite-dimensional $\mathbb{F}$-algebra, we say that a homomorphism of algebras T from $\mathcal{A}$ to $\mathrm{End}_{\mathbb{F}}(V)$ – where V is an n-dimensional $\mathbb{F}$-vector space – is a representation of degree n of $\mathcal{A}$. From this, we can establish a correspondence between the representations of $\mathcal{A}$ and its $\mathcal{A}$-modules. In fact, if (T, V) is a representation of $\mathcal{A}$ over $\mathbb{F}$, then V is an $\mathcal{A}$-module if we define $Av := T(A)v$ for any $A \in \mathcal{A}, v \in V$. On the other hand, if V is any $\mathcal{A}$-module, we can define a representation (T, V) of $\mathcal{A}$ where each $A \in \mathcal{A}$ is mapped to the endomorphism A_L, with $A_L(v) = Av$, that is, the function that applies A to vectors in V. From this correspondence, we can reformulate the concepts introduced earlier in terms of modules: if T is a representation of $\mathcal{A}$, then T is irreducible if V is a simple $\mathcal{A}$-module, and T is completely reducible if V is a semisimple $\mathcal{A}$-module.

This correspondence, together with Proposition 2.2.1, allows us to characterize semisimplicity in terms of representations: an algebra will be semisimple if and only if every representation (T, V) is completely reducible. We can also note that if $\mathcal{A}$ is semisimple over an algebraically closed field $\mathbb{F}$, and if we write it as

$$\mathcal{A} = \prod_{i=1}^{m} \mathcal{A}_i,$$

where $\mathcal{A}_i \cong_{\mathcal{A}} \mathcal{L}_i^{d_i}$, each $\mathcal{L}_i$ is a simple submodule, and $\mathcal{L}_i \ncong_{\mathcal{A}} \mathcal{L}_j$ if $i \neq j$, then every irreducible representation (T, V) will be of the form $V \cong_{\mathcal{A}} \mathbb{F}^{d_i}$ for some d_i. In fact, since V is a simple $\mathcal{A}$-module, it follows from Proposition 2.2.3 that it must be isomorphic as an $\mathcal{A}$-module to one of the direct summands $\mathcal{L}_i$ of $\mathcal{A}$. From Wedderburn's Theorem, we know that $\mathcal{A}_i$ and $M_{d_i}(\mathbb{F})$ are isomorphic algebras for any i, so each $\mathcal{L}_i$ is isomorphic to a simple submodule of some $M_{d_i}(\mathbb{F})$, and by Example 1.3.14, these submodules are all isomorphic to $\mathbb{F}^{d_i}$, so $V \cong_{\mathcal{A}} \mathcal{L}_i \cong_{\mathcal{A}} \mathbb{F}^{d_i}$.

With these results, we can now return to the case of finite groups. First, we note that the representations of G correspond to the representations of $\mathbb{F}G$. In fact, if (T, V) is a representation of G, define the following map:

$$T' : \mathbb{F}G \mapsto \mathrm{End}_{\mathbb{F}}(V),$$
$$T'(\sum_{g \in G} \alpha_g g) = \sum_{g \in G} \alpha_g T(g),$$

then (T', V) is a representation of $\mathbb{F}G$. Similarly, if (T', V) is a representation of $\mathbb{F}G$, we can define

$$T : G \mapsto \mathrm{GL}(V),$$
$$T(g) = T'(g),$$

then (T, V) is a representation of G, i.e., there is a correspondence between the representations of the group G and its group algebra. Furthermore, it also holds that W is a G-subspace if and only if it is an $\mathbb{F}G$-subspace, so a representation (T, V) of G is completely reducible if and only if the corresponding representation (T', V) of $\mathbb{F}G$ is completely reducible. We can now combine these observations with Maschke's Theorem to obtain the following result:

Corollary 2.5.5. If G is a finite group and $\mathbb{F}$ is a field whose characteristic does not divide $|G|$, then $\mathbb{F}G$ is a semisimple algebra.

■

If G and $\mathbb{F}$ satisfy the conditions of the previous result – and additionally, if $\mathbb{F}$ is algebraically closed –, we can then use Wedderburn's Theorem to decompose the group algebra as

$$\mathbb{F}G = \prod_{i=1}^{m} \mathcal{A}_i \cong \prod_{i=1}^{m} M_{d_i}(\mathbb{F}),$$

where $\mathcal{A}_i$ are the components of the Wedderburn decomposition of the algebra, and since the dimension of $\mathbb{F}G$ is precisely $|G|$, it follows that

$$|G| = \sum_{i=1}^{m} d_i^2.$$

We say that a function $\varphi : G \mapsto \mathbb{F}$ is a *class function* of G if $\varphi(g) = \varphi(hgh^{-1})$, for any $g, h \in G$, i.e., φ is constant on the conjugacy classes of G. We can show that the set of class functions can be identified with the center $Z(\mathbb{F}G)$:

Theorem 2.5.6. If G is a finite group and $\mathbb{F}$ is an algebraically closed field, then the dimension of $Z(\mathbb{F}G)$ is equal to the number of conjugacy classes of G, and every class function $\varphi : G \mapsto \mathbb{F}$ can be identified with $\sum_{g \in G} \varphi(g) g \in Z(\mathbb{F}G)$. In particular, the dimension of $Z(\mathbb{F}G)$ is equal to the number of non-isomorphic irreducible representations of G over $\mathbb{F}$.

Proof. Let $\mathcal{C}_1, \ldots, \mathcal{C}_s$ be the conjugacy classes of G, and consider the elements

$$C_i = \sum_{g \in \mathcal{C}_i} g \in \mathbb{F}G.$$

Since the conjugacy classes partition G, it follows that $C_1, \ldots, C_s$ are linearly independent, and if $h \in G$, then $hC_ih^{-1} = C_i$, so each $C_i \in Z(\mathbb{F}G)$, and thus the set of class functions is certainly contained in $Z(\mathbb{F}G)$. To see the reverse inclusion, take an element $x = \sum_{g\in G} \alpha_g g \in Z(\mathbb{F}G)$, and we have that

$$\sum_{g\in G} \alpha_g g = h^{-1}(\sum_{g\in G} \alpha_g g)h = \sum_{g\in G} \alpha_g h^{-1}gh = \sum_{g\in G} \alpha_{hgh^{-1}} g,$$

where the last equality follows from the change of variable $t = h^{-1}gh$, so $\alpha_g = \alpha_{hgh^{-1}}$ for any $h \in G$, and thus x is constant on the conjugacy classes of G. This shows that the class functions can be identified with the elements of $Z(\mathbb{F}G)$, which has dimension s, and we have previously shown that the dimension of the center of a semisimple algebra is exactly the number of isomorphism classes of simple modules, which in our case are precisely the irreducible representations of G over $\mathbb{F}$. ■

Example 2.5.7. We can consider the symmetric group S_3 with 6 elements, and its group algebra $\mathbb{C}S_3$ of dimension 6. This algebra is semisimple, so its dimension must be a sum of squares, but there are only two possibilities for writing 6 as a sum of squares:

$$6 = 1 + 1 + 1 + 1 + 1 + 1 \quad \text{and} \quad 6 = 1 + 1 + 2^2.$$

Since S_3 is not abelian, it follows that $\mathbb{C}S_3$ is not a commutative algebra, and thus we are left with only the second possibility. Therefore,

$$\mathbb{C}S_3 \cong \mathbb{C} \times \mathbb{C} \times M_2(\mathbb{C}),$$

that is, there are 3 isomorphism classes of irreducible representations of S_3, two of degree 1 and one of degree 2. The center of $\mathbb{C}S_3$ will have dimension 3, which is precisely the number of conjugacy classes of S_3.

3 Matrix Algebras

We will now study subsets of $M_n(\mathbb{C})$ that form algebras, called *matrix algebras*. These algebras are of great importance in various areas of mathematics, and are particularly interesting for computational applications.

We denote the identity matrix in $M_n(\mathbb{C})$ by I, and the matrix with all elements equal to 1 as J. If n is a natural number, we define the set $[n] := \{1, \ldots, n\}$ as all the natural numbers up to n. We remind the reader that if V is a vector space over $\mathbb{C}$, and if there exists a function

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

that is linear in the first variable, satisfies $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$, and $\langle v, v \rangle$ is real and positive if $v \neq 0$, then we say that V is an *inner product space*. Such a function determines an *induced norm* $||v|| := \langle v, v \rangle^{1/2}$ that maps elements of V to real numbers. If V, W are two vector spaces over $\mathbb{C}$ with inner products $\langle \cdot, \cdot \rangle_1, \langle \cdot, \cdot \rangle_2$, respectively, and if φ is a $\mathbb{C}$-homomorphism between V and W, we define the *adjoint* φ^* of φ as the $\mathbb{C}$-homomorphism between W and V such that for any $v \in V$ and $u \in W$ we have

$$\langle v, \varphi^*(u) \rangle_1 = \langle \varphi(v), u \rangle_2,$$

and in the case where $W = V$ and $\varphi = \varphi^*$, we say that φ is *self-adjoint*. If W is a subspace of V, we define

$$W^\perp := \{v \in V \mid \forall w \in W : \langle v, w \rangle = 0\}$$

as the *orthogonal subspace* of W, and if V has finite dimension, we can always decompose

$$V = W \oplus W^\perp$$

as a direct sum of vector spaces. In this case, there exist unique self-adjoint and idempotent operators $P_W, P_{W^\perp}$ – called orthogonal projections – that map a given matrix X to its respective component in W or $W^\perp$.

In the case of matrices, the adjoint of a matrix $A \in M_n(\mathbb{C})$ is simply the conjugate-transpose matrix $A^* = (\overline{A})^T$. We can define an inner product for the matrix algebra $M_n(\mathbb{C})$ as follows:

$$\langle A, B \rangle := \operatorname{tr}(AB^*),$$

where tr denotes the *trace* of a matrix. We leave it to the reader to verify that the above function is indeed an inner product, and it is also worth noting that for any matrix $A \in M_n(\mathbb{C})$, the matrix AA^* is Hermitian, i.e., has real eigenvalues, and therefore its trace is also a real number.

We can also formalize two operations that will be quite common from now on: the operation of mapping a vector $v = (v_1, \ldots, v_n)$ to a diagonal matrix $\operatorname{Diag}(v)$ with diagonal entries given by $v_1, \ldots, v_n$, and the operation of mapping any matrix A to the vector $\operatorname{diag}(A) = (A_{11}, \ldots, A_{nn})$ of its diagonal entries. Formally, we define the function

$$\operatorname{Diag} : \mathbb{C}^n \mapsto M_n(\mathbb{C}),$$

where

$$\operatorname{Diag}(v) = \begin{pmatrix} v_1 & 0 & \ldots & 0 \\ 0 & v_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & v_n \end{pmatrix},$$

and similarly, we define the function

$$\begin{aligned} \operatorname{diag} : M_n(\mathbb{C}) &\mapsto \mathbb{C}^n, \\ \operatorname{diag}(A) &= (A_{11}, \ldots, A_{nn}). \end{aligned}$$

Note that both are $\mathbb{C}$-linear functions, and that $\operatorname{Diag}^* = \operatorname{diag}$, i.e., one is the adjoint of the other.

We recall that a Hermitian matrix $X \in M_n(\mathbb{C})$ is said to be *positive semidefinite* (PSD) if, for any vector $v \in \mathbb{C}^n$, we have $v^*Xv \geq 0$. Equivalently, X has only real non-negative eigenvalues, and therefore if we write the diagonalization $X = U\Lambda U^*$, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues and U is a unitary matrix of eigenvectors, we can define

$$X^{1/2} := U\text{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})U^*.$$

Note that $X^{1/2}$ is clearly Hermitian, and in this case, it easily follows that X is PSD if, and only if, there exists a matrix B such that $X = BB^*$, because we can take $B = U\text{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. We denote the set of symmetric $n \times n$ matrices with real entries by $\mathbb{S}^n$, the subset of $\mathbb{S}^n$ formed by PSD matrices with real entries will be denoted by $\mathbb{S}^n_+$, and the subset of positive definite matrices with real entries will be denoted by $\mathbb{S}^n_{++}$. The set $\mathbb{S}^n_+ \subseteq \mathbb{S}^n$ is not a vector subspace because, although it is closed under addition, it is not closed under scalar multiplication, but it is a *cone* over $\mathbb{R}$, i.e., it is a closed subset under multiplication by non-negative real numbers, and it is also a *convex* set.

3.1 $*$-Algebras

Let $\mathcal{A}$ be a matrix algebra in $M_n(\mathbb{C})$, that is, a $\mathbb{C}$-subspace closed under matrix multiplication that has a multiplicative identity element. We say that $\mathcal{A}$ is a $*$-*algebra* – or a *self-adjoint algebra* – if for any element A in $\mathcal{A}$, its conjugate-transpose A^* also belongs to $\mathcal{A}$. A subalgebra $\mathcal{B}$ of $\mathcal{A}$ that is also closed under the conjugate-transpose is called a $*$-*subalgebra*, and a homomorphism φ of algebras is called a $*$-*homomorphism* if $\varphi(A^*) = \varphi(A)^*$.

Example 3.1.1 (Polynomial Algebra)**.** If we consider any Hermitian matrix $A \in M_n(\mathbb{C})$, the set $\mathbb{C}[A]$ of all polynomials in A, that is, expressions of the form

$$a_0I + a_1A + a_2A^2 + ... + a_kA^k,$$

is a commutative $*$-algebra with identity I. If $A = \sum_{i=1}^d \lambda_i E_i$ is the spectral decomposition of A, where d is the number of distinct eigenvalues, we have that

$$E_i = \prod_{j \neq i} \frac{A - \lambda_j I}{\lambda_i - \lambda_j},$$

that is, E_i is a polynomial in A, and therefore $E_i \in \mathbb{C}[A]$. It follows that the dimension of $\mathbb{C}[A]$ is exactly d, since the matrices $\{E_i\}$ form a linearly independent set, and $A^k = \sum_{i=1}^d \lambda_i^k E_i$ for any k, so they generate $\mathbb{C}[A]$. We can then decompose $\mathbb{C}[A]$ as

$$\mathbb{C}[A] = \mathbb{C}[A]E_1 \oplus \ldots \oplus \mathbb{C}[A]E_d = \mathbb{C}E_1 \oplus \ldots \oplus \mathbb{C}E_d,$$

where the matrices E_i will be the primitive central idempotents of the algebra, and therefore $\mathcal{A}$ is semisimple. Each component $\mathbb{C}E_i$ is a simple $\mathbb{C}[A]$-module of dimension 1, and since all the matrices in $\mathbb{C}[A]$ are polynomials in A, we can consider the unitary matrix U with the normalized eigenvectors of A in its columns so that $U^*\mathbb{C}[A]U$ is a set of diagonal matrices. It is also interesting to consider the centralizer of A in $M_n(\mathbb{C})$, that is, the set

$$C(A) = \{B \in M_n(\mathbb{C}) \mid [A, B] = 0\},$$

and note that C is a matrix algebra with identity I that contains $\mathbb{C}[A]$ as a subalgebra, so the dimension of $C(A)$ is at least $d + 1$. Furthermore, $C(A)$ will be a $*$-algebra, because if $B \in C(A)$, we have

$$B^*A = (A^*B)^* = (AB)^* = (BA)^* = AB^*.$$

Example 3.1.2 (Centralizers of Permutation Groups)**.** An important example of $*$-algebras is the centralizers of permutation groups. Formally, a permutation group is nothing more than a subgroup of the symmetric group S_n, and there is a natural correspondence between permutation groups and subgroups of $GL_n(\mathbb{C})$

formed by permutation matrices. If $G \subseteq GL_n(\mathbb{C})$ is a subgroup formed by permutation matrices, its centralizer $C(G)$ will be a matrix algebra with identity I. If $A \in C(G), P \in G$, noting that $P^T = P^*$ and that $P^T \in G$, we have

$$A^*P = (P^TA)^* = (AP^T)^* = PA^*,$$

so $C(G)$ is a $*$-algebra. A case that will be particularly interesting for the upcoming chapters is when we consider the centralizer of the automorphism group of a graph X, that is, the subgroup of $GL_n(\mathbb{C})$ formed by the permutation matrices P such that $PAP^T = A$, where A is the adjacency matrix of X. The fact that this set forms a $*$-algebra will be useful for inferring combinatorial properties of certain graphs that exhibit high regularity.

3.2 Triangularization and Diagonalization of Commutative Algebras

This section focuses on results related to commutative algebras, which will later serve as the basis for proving facts about $*$-algebras. First, we will show that every commutative matrix algebra in $M_n(\mathbb{C})$ can be simultaneously triangularized, that is, there exists an orthogonal basis where all the matrices of the algebra are in upper triangular form. From there, we will conclude that every commutative $*$-algebra can be simultaneously diagonalized, meaning there exists a common eigenvector basis for all the matrices in the algebra.

Before we show the general result, we will need two auxiliary results, which are proven below:

Lemma 3.2.1. Let A be a matrix in $M_n(\mathbb{C})$, and W a A-invariant subspace of $\mathbb{C}^n$. Then there exists an eigenvector of A in W.

Proof. Let $B = (v_1 \quad v_2 \quad \dots \quad v_k)$ be an $n \times k$ matrix with a basis for W in its columns, where k is the dimension of W. The fact that W is A-invariant is equivalent to saying that

$$AB = BC,$$

where C is some $k \times k$ matrix with the coefficients of the action of A on each of the elements of the basis of W. Since $C \in M_k(\mathbb{C})$, we can take an eigenvector $v \in \mathbb{C}^k$ with eigenvalue λ for C, so

$$ABv = BCv = \lambda Bv,$$

thus Bv is an eigenvector of A with eigenvalue λ. ■

Lemma 3.2.2. If $\{A_1, ..., A_d\}$ is a set of matrices in $M_n(\mathbb{C})$ that commute with each other, then there exists a vector v that is a common eigenvector for all the matrices.

Proof. We will prove this by induction on d. In the base case, note that since A_1 and A_2 commute, we have that if v is an eigenvector of A_1 with eigenvalue λ, then for any k

$$A_1(A_2^kv) = A_2^k(A_1v) = \lambda(A_2^kv),$$

so A_2^kv is an eigenvector of A_1. Therefore, we consider the space

$$W = \text{span}_{\mathbb{C}}(v, A_2v, A_2^2v, \dots),$$

which by construction is A_2-invariant, so by the previous lemma it contains an eigenvector of A_2, but since every element of W is also an eigenvector of A_1, we obtain a common eigenvector. For the general case, we use induction to find a common eigenvector for $A_1, ..., A_{d-1}$, and then apply the same reasoning in

$$W = \text{span}_{\mathbb{C}}(v, A_dv, A_d^2v, \dots),$$

thus obtaining a common eigenvector for $A_1, ..., A_d$. ■

Now we are ready to prove the main result of this section.

Theorem 3.2.3. Every commutative matrix algebra $\mathcal{A}$ in $M_n(\mathbb{C})$ can be simultaneously triangularized by an orthogonal basis. In other words, there exists a unitary matrix U such that the matrices of $U^*\mathcal{A}U$ are all in upper triangular form.

Proof. Let $\{A_1, ..., A_d\}$ be a basis for $\mathcal{A}$. We will prove the result by induction on n. The base case is immediate, and for the general case we first find a common eigenvector v_1 for $A_1, ..., A_d$ using the previous lemma, and then decompose the space as

$$\mathbb{C}^n = \mathbb{C}v_1 \oplus W,$$

where $W = (\mathbb{C}v_1)^\perp$. Note then that each matrix A_i can be written with respect to a basis formed by v_1 and a basis for W, and such a matrix will be of the form

$$A_i = \begin{pmatrix} \lambda_i & a_i \\ 0 & A_i|_W \end{pmatrix},$$

where λ_i is the eigenvalue of A_i associated with v_1, a_i is a row vector of dimension $n-1$. Since $\mathcal{A}$ is commutative, it follows that the vector space generated by the matrices $\{A_i|_W\}$ is a commutative matrix algebra in $M_{n-1}(\mathbb{C})$, so by induction we can find a unitary matrix U such that $U^*A_i|_W U$ is upper triangular. From this it follows that the change of basis given by

$$\begin{pmatrix} 1 & 0 \\ 0 & U^* \end{pmatrix} \begin{pmatrix} \lambda & a \\ 0 & A_i|_W \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & U \end{pmatrix} = \begin{pmatrix} \lambda & a' \\ 0 & U^*A_i|_W U \end{pmatrix}$$

puts all the matrices A_i in upper triangular form, so $\mathcal{A}$ can be simultaneously triangularized by an orthogonal basis, as we wanted. ■

We note that the previous result implies that every matrix in $M_n(\mathbb{C})$ can be triangularized by an orthogonal basis. It is also common to find versions of the previous result stated in terms of a commutative family of matrices, that is, an arbitrary set $\{A_i\}$ of commuting matrices in $M_n(\mathbb{C})$, but note that the result at hand immediately implies these other versions: simply consider the algebra generated by the family, which will be a subalgebra of $M_n(\mathbb{C})$ of finite dimension, and then apply our result to any basis. In the particular case of commutative $*$-algebras, we obtain the following corollary:

Corollary 3.2.4. Every commutative $*$-algebra $\mathcal{A}$ of matrices in $M_n(\mathbb{C})$ has a common orthogonal eigenvector basis. In other words, there exists a unitary matrix U such that the matrices of $U^*\mathcal{A}U$ are all in diagonal form.

Proof. First, note that if W is an A-invariant subspace, then $W^\perp$ is A^*-invariant. If $W = \mathbb{C}v$ where v is an eigenvector of A, then since A is normal, it follows that v is also an eigenvector of A^*, so W is A^*-invariant, implying that $W^\perp$ is A-invariant. Therefore, in the proof of the previous Theorem, when we take a common eigenvector v_1 for a basis of $\mathcal{A}$, the space $W = (\mathbb{C}v_1)^\perp$ will be invariant under the matrices of the basis, and then we can proceed inductively to obtain a simultaneous diagonalization of $\mathcal{A}$ by a common orthogonal eigenvector basis U. ■

3.3 Semisimplicity of $*$-Algebras

Now we are ready to present the elementary proof of the main result of this chapter: every $*$-algebra is semisimple. First, we will demonstrate a very useful auxiliary result: every matrix algebra with respect to the Schur product has a unique basis of 01-matrices – that is, matrices with entries in $\{0, 1\}$ – which is orthogonal with respect to the Schur product. We will prove the result in two ways: the first follows almost immediately from the characterizations of semisimplicity seen in the previous chapter, while the second is elementary and constructs the basis explicitly.

Lemma 3.3.1. If $\mathcal{A} \subseteq M_n(\mathbb{C})$ is a matrix algebra with respect to the Schur product, then there exists a unique basis of matrices $A_1, ..., A_d$ with entries in $\{0, 1\}$ for $\mathcal{A}$ such that $A_i \circ A_j = 0$ if $i \neq j$.

Proof 1. Let $\mathcal{A}$ be a matrix algebra with respect to the Schur product. Denote by $A^{\circ k}$ the Schur product of a matrix A with itself k times, and note that $A_{ij}^{\circ k} = A_{ij}^k$. Therefore, if $A \in \mathcal{A}$ is a nilpotent element, it follows that $A = 0$, and so the Jacobson radical of $\mathcal{A}$ is $\{0\}$, meaning that $\mathcal{A}$ is semisimple. This shows that every matrix algebra with respect to the Schur product is semisimple, and from this it follows that we can consider its central primitive orthogonal idempotents $A_1, ..., A_d$ such that

$$E = A_1 + \ldots + A_d.$$

Note that the matrices A_i are idempotent with respect to the Schur product, so $A_i \circ A_i = A_i$, meaning that A_i is 01, and also that $A_i \circ A_j = 0$ if $i \neq j$. Since every matrix algebra with respect to the Schur product is commutative, by Theorem 2.4.2 it follows that the idempotents $A_1, ..., A_d$ form a basis for $\mathcal{A}$, and the uniqueness of this basis follows from the uniqueness of the idempotents. ■

Proof 2. To demonstrate the result in an elementary way, we will first make two observations about polynomials in matrices with respect to the Schur product. Let $A \in M_n(\mathbb{C})$ be any matrix, and write

$$A = \sum_{i=1}^{l} \alpha_l A_l,$$

where $\{\alpha_1, ..., \alpha_l\}$ is the set of distinct non-zero entries of A, and the matrices A_l are the orthogonal 01 components of A with entries equal to 1 in the positions where α_l occurs in A, and zero elsewhere, with $A_i \circ A_j = 0$ if $i \neq j$, and $A \circ A_i = \alpha_i A_i$. Now, if $p(x) = \sum_{i=0}^{n} a_n x^n$ is a polynomial with coefficients in $\mathbb{C}$, we can define the expression

$$p \circ A = a_0 E + a_1 A + \cdots + a_n A^{\circ n},$$

and note that

$$p \circ A = \sum_{i=1}^{l} p(\alpha_l) A_l,$$

because since A_l is a 01 matrix, it follows that it is idempotent with respect to the Schur product, meaning that $A_l^{\circ k} = A_l$ for any k. In particular, we can define the polynomials[5]

$$p_i(x) = \prod_{j \neq i} \frac{x - \alpha_j}{\alpha_i - \alpha_j},$$

such that $p_i(\alpha_i) = 1$ and $p_i(\alpha_j) = 0$ if $j \neq i$, so $p_i(A) = A_i$, and thus it is always possible to find a polynomial in terms of the Schur product that maps A to one of its 01 components. Since $\mathcal{A}$ is closed under the Schur product, this means that it contains all the 01 components of its matrices, so $\mathcal{A}$ is generated by a set of 01 matrices. We can then consider the primitive elements, that is, those 01 components that cannot be decomposed into a sum of two or more 01 components in $\mathcal{A}$, and note that these components form a set of orthogonal idempotents with respect to the Schur product that generate $\mathcal{A}$. Moreover, this set must be linearly independent due to its orthogonality, so it forms the desired basis for $\mathcal{A}$. Any other orthogonal idempotent basis must be formed by primitive elements, so it will be a permutation of the previously found basis. ■

The basis of orthogonal 01 matrices for the Schur product found in the last proof is known as the *Schur basis* of an algebra with respect to the Schur product. With this result, we are now able to show that every commutative $*$-algebra is semisimple.

Theorem 3.3.2. Every commutative $*$-algebra is semisimple. In particular, if $\mathcal{A}$ is a commutative $*$-algebra of dimension d, then there exists a unitary matrix U and a partition $[n] = S_0 \sqcup S_1 \sqcup ... \sqcup S_d$, with $S_1, ..., S_d$ non-empty, such that

$$\mathcal{A} = U\{\lambda_1 I_1 + ... + \lambda_d I_d \mid \lambda_i \in \mathbb{C}\}U^*,$$

where each I_i is the diagonal matrix with ones in the positions of S_i and zeros elsewhere.

[5]The polynomials in question are known as *Lagrange interpolating polynomials*.

Proof. Note that if $A \in \mathcal{A}$, then $A^* \in \mathcal{A}$ by definition, and since the algebra is commutative, it follows that A is normal. If we fix a basis for $\mathcal{A}$, this implies that such a basis consists of normal matrices that commute, so Corollary 3.2.4 guarantees the existence of a matrix U that simultaneously diagonalizes all matrices in the basis, and consequently all matrices in $\mathcal{A}$. A diagonal matrix algebra is closed under the Schur product, so we can apply Lemma 3.3.1 to obtain a basis $I_1, ..., I_d$ of 01 diagonal matrices with disjoint support, so we define S_i as the set of indices of the respective non-zero entries of I_i, and S_0 as the set of indices in $[n]$ that correspond to zero entries in all elements of the basis, i.e., $I_0 = I - (I_1 + ... + I_d)$. Thus we obtain a partition $[n] = S_0 \sqcup S_1 \sqcup ... \sqcup S_d$, and every element of $\mathcal{A}$ will be a linear combination of $I_1, ..., I_d$. ■

The matrices $E_1, ..., E_d$ given by $E_i = UI_iU^*$ obtained from the matrices in the previous theorem are precisely the centrally primitive orthogonal idempotents of $\mathcal{A}$, because $E_iE_j = \delta_{ij}E_i$, $E_1 + ... + E_d$ is the identity, and the dimension of the algebra is exactly d. Furthermore, each E_i is an orthogonal projector onto a sub-eigenspace, i.e., a subspace contained in an eigenspace, of each matrix in $\mathcal{A}$, and in particular E_i is PSD. Note also that $S_0 = \emptyset$ implies that $I \in \mathcal{A}$, and in this case the algebra $\mathcal{A}$ will be a subalgebra of $M_n(\mathbb{C})$. Since these idempotents form a basis for $\mathcal{A}$, it follows that every matrix A in the algebra can be written as $A = \sum_{i=1}^{d} \alpha_i E_i$ and that $AE_i = \alpha_i E_i$, so the Wedderburn decomposition of $\mathcal{A}$ is given by

$$\mathcal{A} = \bigoplus_{i=1}^{d} \mathcal{A}E_i = \bigoplus_{i=1}^{d} \mathbb{C}E_i = \prod_{i=1}^{d} \mathbb{C}E_i.$$

With this, we are ready to prove the main theorem of this section. Similarly to what was done earlier, we will present more than one proof for the result. The first is a constructive and elementary proof due to [BGSV12], and the second and third are proofs that strongly depend on the concepts developed in the previous chapter about semisimple algebras and their various characterizations.

Theorem 3.3.3. Every $*$-algebra $\mathcal{A}$ in $M_n(\mathbb{C})$ is semisimple. In particular, we can write

$$\mathcal{A} = \bigoplus_{i=1}^{m} \mathcal{A}E_i,$$

where the matrices E_i are the centrally primitive orthogonal idempotents of $\mathcal{A}$.

Proof 1 ([BGSV12, Theorem 9.1]). To show the result, we will follow this strategy: first, we will consider a maximal commutative $*$-subalgebra contained in $\mathcal{A}$ and obtain its minimal idempotents, then we will construct an equivalence relation from this set of idempotents, and finally, we will obtain the centrally primitive orthogonal idempotents of $\mathcal{A}$ from the equivalence classes of this relation.

We start with $\mathcal{B} \subseteq \mathcal{A}$ being a maximal commutative $*$-subalgebra—and note that $Z(\mathcal{A}) \subseteq \mathcal{B}$. First, we show that $C_{\mathcal{A}}(\mathcal{B}) \subseteq \mathcal{B}$, i.e., $\mathcal{B}$ contains all matrices that commute with all of its elements. Indeed, let $A \in C_{\mathcal{A}}(\mathcal{B})$ be a matrix in the centralizer of the subalgebra, and consider the following possibilities:

(i) If A is normal, note that if $A \notin \mathcal{B}$, then the algebra generated by $\mathcal{B} \cup \{A, A^*\}$ is a commutative $*$-subalgebra strictly containing $\mathcal{B}$, contradicting its maximality;

(ii) If A is not normal, we can consider the matrix $A + A^*$, which is certainly normal, and therefore by the same argument as in the previous item, $A + A^*$ must belong to $\mathcal{B}$. But since $A \in C_{\mathcal{A}}(\mathcal{B})$, we would have $A(A + A^*) = (A + A^*)A$, leading to a contradiction, as this implies that A is normal.

Thus, we conclude that $\mathcal{B}$ contains its centralizer. Using Theorem 3.3.2, we can obtain a unitary matrix U that, after replacing $\mathcal{A}$ with $U^*\mathcal{A}U$, allows us to assume without loss of generality that $\mathcal{B}$ is in diagonal form, with minimal idempotents given by the matrices $I_0, I_1, ..., I_d$ with 01 entries that form an orthogonal basis for $\mathcal{B}$, and with partition $[n] = S_0 \sqcup S_1 \sqcup ... \sqcup S_d$. Since the identity in $\mathcal{B}$ is the same as in $\mathcal{A}$, it follows that $I_1 + ... + I_d$ is the identity of $\mathcal{A}$, but note that in general the matrices I_i do not need to belong to the center of $\mathcal{A}$, simply because $Z(\mathcal{A}) \subseteq Z(\mathcal{B})$, meaning the matrices I_i are still not the centrally primitive orthogonal idempotents of $\mathcal{A}$, and to obtain them, we need a bit more effort.

Now we will fix $A \in \mathcal{A}$ and define the matrices A_{ij}, with $i, j \in \{0, ..., d\}$ and size $|S_i| \times |S_j|$, given by the restrictions of I_iAI_j to the rows of S_i and columns of S_j. We observe that:

(i) $A_{00} = 0$, and A_{ii} is a multiple of the identity matrix in $M_{|S_i|}(\mathbb{C})$. To see this, first note that the matrix $I_i A I_i$ is a diagonal matrix with non-zero entries given by the diagonal entries of A indexed by S_i, and also that if $i \neq j$,

$$(I_i A I_i) I_j = 0 = I_j (I_i A I_i),$$

so $I_i A I_i$ commutes with the basis of $\mathcal{B}$, and hence $I_i A I_i \in C_{\mathcal{A}}(\mathcal{B}) \subseteq \mathcal{B}$. In the case of $I_0 A I_0$, since $I_0 \mathcal{B} I_0 = \{0\}$, we obtain $I_0 A I_0 = 0$. For the remaining i, since the basis is formed by diagonal matrices with disjoint support, we have that the coefficients of $I_i A I_i$ with respect to the basis $I_1, ..., I_d$ will be zero for any $j \neq i$, i.e., $I_i A I_i$ is indeed a multiple of I_i, and its restriction in $M_{|S_i|}(\mathbb{C})$ will be a multiple of the identity matrix;

(ii) A_{ij} is either zero or a positive multiple of a unitary matrix, and in this case, the cardinalities of S_i are equal. Indeed, take A_{ij} non-zero and assume without loss of generality that $|S_i| \geq |S_j|$, and note that

$$(I_i A I_j)(I_i A I_j)^* = I_i (A I_j A^*) I_i,$$

so by (i) it follows that $(A I_j A^*)_{ii}$ is a multiple of the identity in $M_{|S_i|}(\mathbb{C})$, implying that $A_{ij} A_{ij}^*$ is also, so $\mathrm{rk}(A_{ij}) = |S_i|$, but on the other hand, $\mathrm{rk}(A_{ij}) \leq |S_j|$, so $|S_i| = |S_j|$, and hence the matrices A_{ij} are square for all i, j. This implies that $A_{ij} A_{ij}^*$ is a PSD matrix, and must therefore be a positive real multiple—denoted by α—of the identity in $M_{|S_i|}(\mathbb{C})$, and then $\sqrt{\alpha} A_{ij}$ is a unitary matrix, making A_{ij} a positive multiple of a unitary matrix.

With these observations, we proceed to the final part of the proof. We define a relation $\sim$ on the set $[d]$ as follows: $i \sim j$ if and only if $I_i \mathcal{A} I_j \neq \{0\}$, i.e., index i is related to index j if and only if there exists a matrix A in $\mathcal{A}$ such that the matrix formed by the rows of A indexed by S_i and the columns of A indexed by S_j is not entirely zero. The relation is reflexive by observation (i), and it is also symmetric because $\mathcal{A}$ is a $*$-algebra, so $I_i \mathcal{A} I_j = (I_j \mathcal{A} I_i)^*$. The transitivity follows from observation (ii): if $i \sim j$ and $j \sim k$, then there exist matrices $A, B \in \mathcal{A}$ such that $I_i A I_j$ and $I_j B I_k$ are non-zero, and therefore A_{ij}, B_{jk} are positive multiples of unitary matrices of the same size (since $|S_i| = |S_j| = |S_k|$), and thus the product $A_{ij} B_{jk}$ is a unitary matrix, implying that

$$0 \neq (I_i A I_j)(I_j B I_k) = I_i (A I_j B) I_k \in I_i \mathcal{A} I_k,$$

so $i \sim k$.

Now, consider $\{E_1, ..., E_m\} = \{\sum_{j \sim i} I_j \mid i \in [d]\}$ as the 01 diagonal matrices induced by the equivalence classes $\sim$ on the minimal idempotents of $\mathcal{B}$. We claim that the matrices E_i are the centrally primitive orthogonal idempotents of $\mathcal{A}$. Indeed, first note that by construction each E_i is a Hermitian and idempotent matrix, $E_1 + ... + E_m$ is the identity in $\mathcal{A}$, and if $i \neq j$, then $E_i \mathcal{A} E_j = \{0\}$, so the idempotents are orthogonal. If $A \in \mathcal{A}$, then

$$\begin{aligned}
AE_i &= (\sum_j E_j) A E_i \\
&= \sum_j E_j A E_i \\
&= E_i A E_i \\
&= E_i A E_i + \sum_{j \neq i} E_i A E_j \\
&= \sum_j E_i A E_j \\
&= E_i A (\sum_j E_j) \\
&= E_i A,
\end{aligned}$$

implying that each $E_i \in Z(\mathcal{A})$. We can then decompose $\mathcal{A}$ as

$$\mathcal{A} = \bigoplus_{i=1}^{m} \mathcal{A}E_i = \prod_{i=1}^{m} \mathcal{A}E_i$$

where the sets $\mathcal{A}E_i$ are $\mathcal{A}$-submodules that are semisimple by construction. The sum is direct because the submodules are written in terms of orthogonal idempotents, thus $\mathcal{A}$ is semisimple, as we wanted. ■

The proof in question is interesting for several reasons, but perhaps the main one is the fact that it is a constructive proof. It is possible to find a maximal commutative $*$-subalgebra in linear time in the dimension of the algebra $\mathcal{A}$, that is, the proof gives us an efficient algorithm to find the idempotents of the algebra, and from these idempotents, it is possible to obtain a common block-diagonal basis for all the matrices of $\mathcal{A}$. Now we will proceed with the other proofs.

Proof 2. Let $\mathcal{A}$ be a $*$-algebra, and let $\mathcal{B} \subseteq \mathcal{A}$ be an $\mathcal{A}$-submodule. We know that $\mathcal{B}$ is also a vector subspace of $\mathcal{A}$, so we can decompose

$$\mathcal{A} = \mathcal{B} \oplus \mathcal{B}^{\perp},$$

where $\mathcal{B}^{\perp}$ is the orthogonal complement of $\mathcal{B}$ in $\mathcal{A}$. Fix $X \in \mathcal{B}^{\perp}$, and take any element $A \in \mathcal{A}$, so for any $B \in \mathcal{B}$ we have:

$$\begin{aligned}\langle AX, B\rangle &= \operatorname{tr}(AXB^*) \\ &= \operatorname{tr}(XB^*A) \\ &= \operatorname{tr}(X(A^*B)^*) \\ &= \langle X, A^*B\rangle.\end{aligned}$$

Since $A \in \mathcal{A}$, it follows that $A^* \in \mathcal{A}$, and since $\mathcal{B}$ is an $\mathcal{A}$-submodule, it follows that $A^*B \in \mathcal{B}$. Therefore $\langle X, A^*B\rangle = 0$, and from this we have that $\langle AX, B\rangle = 0$, i.e., $\mathcal{B}^{\perp}$ is also an $\mathcal{A}$-submodule. Thus, we conclude that every $\mathcal{A}$-submodule of $\mathcal{A}$ is a direct sum, and by Theorem 2.1.2, it follows that $\mathcal{A}$ is semisimple. ■

Proof 3. Let $\mathcal{A}$ be a $*$-algebra, and take an element $A \in J(\mathcal{A})$ in its Jacobson radical. Since the radical is an ideal, we have that $A^*A \in J(\mathcal{A})$, and from Proposition 2.3.2 we know that this implies there exists a natural number r such that $(A^*A)^r = 0$. On the other hand, note that A^*A is a Hermitian matrix, so there exist orthogonal projectors E_θ for every eigenvalue θ of A^*A such that

$$A^*A = \sum_{\theta} \theta E_\theta$$

thus

$$(A^*A)^r = \sum_{\theta} \theta^r E_\theta.$$

In particular, if we take a non-zero eigenvector v associated with any eigenvalue θ, we have

$$0 = (A^*A)^r v = \theta^r v,$$

but this occurs if and only if $\theta = 0$, implying that $A^*A = 0$. Now take a non-zero vector v, and note that $(Av)^*(Av) = 0$ if and only if $Av = 0$, but $(Av)^*(Av) = v^*(A^*A)v = 0$, so $Av = 0$ for any vector in the space, and therefore $A = 0$. This implies that the Jacobson radical of $\mathcal{A}$ is trivial, and therefore by Theorem 2.3.3, it follows that $\mathcal{A}$ is semisimple. ■

4 Association Schemes

Association schemes are mathematical structures that lie at the intersection of algebra and combinatorics. They originally emerged in the study of statistical experiments, but are now widely used in various areas of mathematics and computer science, such as error-correcting code theory and combinatorial design theory. Our goal in this chapter will be to introduce association schemes and also their more general versions, known as coherent configurations, and to show the connections of this theory with the other topics discussed in the previous chapters, with a focus on groups and graphs.

4.1 Configurations and Schemes

4.1.1 Basic Concepts

In this section and throughout the rest of the text, we will discuss relations on finite sets, and we will always assume that any group G is finite. Given a finite set X, a subset $R \subseteq X \times X$ is called a *relation*, that is, R is a set of ordered pairs from X. We can classify a relation R as follows:

- R is *reflexive* if for any $x \in X$, $(x, x) \in R$;
- R is *irreflexive* if for any $x \in X$, $(x, x) \notin R$;
- R is *symmetric* if for any $(x, y) \in R$, $(y, x) \in R$;
- R is *asymmetric* if for any $(x, y) \in X$, $(y, x) \notin R$;
- R is *antisymmetric* if when $(x, y), (y, x) \in R$, then $x = y$;
- R is *transitive* if when $(x, y), (y, z) \in R$, then $(x, z) \in R$.

Whenever we have a relation R, we can consider its *adjacency matrix* $A(R)$, given by a matrix of size $|X| \times |X|$, where

$$A(R)_{xy} = \begin{cases} 1, & \text{if } (x, y) \in R, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, it follows that R is reflexive if and only if $\mathrm{Diag}(A(R)) = I$, and that R is symmetric if and only if $A(R)$ is symmetric. Therefore, we define the relation R^T as the relation whose adjacency matrix is given by $A(R)^T$. Given a relation R, we can fix an element $x \in X$ and consider the set

$$R(x) := \{y \in X | (x, y) \in R\}$$

of neighbors of x in R.

With these observations, we are ready to define coherent configurations.

Definition 4.1.1 (Coherent Configuration)**.** Let X be a finite set, and let $R_0, ..., R_d$ be relations on X with corresponding adjacency matrices $A_0, ..., A_d$. We say that $\{A_0, ..., A_d\}$ – or $(X, \{R_i\}_{i=0}^d)$ – is a *coherent configuration* if the following properties are satisfied:

(1) $\sum_{i=0}^d A_i = J$;

(2) If A_i has a diagonal entry, then A_i is a diagonal matrix;

(3) For each $i \in \{0, ..., d\}$, there exists i' such that $A_i^T = A_{i'}$;

(4) For any $i, j \in \{0, ..., d\}$, there exist non-negative integers p_{ij}^l such that

$$A_i A_j = \sum_{l=0}^{d} p_{ij}^l A_l$$

In terms of relations, item (1) tells us that $R_0, ..., R_d$ partition $X \times X$, (2) tells us that either R_i is irreflexive or it is a relation contained in $\{(x, x)|x \in X\}$, (3) tells us that the set is closed under transposition, and (4) tells us that if we fix elements $x, y \in X$ such that $(x, y) \in R_l$, then the number of neighbors of x in R_i that have y as a neighbor in $R_{j'}$ is a constant p_{ij}^l that depends only on i, j, l, that is, for any $(x, y) \in R_l$, we have

$$p_{ij}^l = |R_i(x) \cap R_{j'}(y)| = |\{z \in X|(x, z) \in R_i, (z, y) \in R_j\}|.$$

Another way to see the above equality is to look at the entry xy of A_iA_j, which will be the product of the x-th row of A_i, which has 1's in the positions z such that $(x, z) \in A_i$ and 0 in the others, with the y-th column of A_j, which has 1's in the positions z such that $(z, y) \in R_j$ – and equivalently $(y, z) \in R'_j$ – and 0 in the others. This product will count the elements z in $R_i(x) \cap R_{j'}(y)$. Combining items (1) and (2), we see that there are diagonal matrices in the configuration with disjoint support whose sum is the identity matrix, and such matrices are called the *fibers* of the configuration. A coherent configuration that has only one fiber, that is, contains I as one of its matrices, is called *homogeneous*, and we define an *association scheme* as a homogeneous coherent configuration. Naturally, if each matrix A_i is symmetric, we say that the configuration is symmetric, and if the matrices A_i commute with respect to matrix multiplication, we say that the configuration is commutative.

4.1.2 Group Configurations

Let us remind the reader that if $X = \{x_1, ..., x_n\}$ is a set with n elements, then the group S_n acts on X by permuting its elements, and we will represent these permutations in cyclic notation, e.g., $(1, 2)$ represents the permutation that maps x_1 to x_2 and vice versa, etc. With this, we can discuss an example of a configuration.

Example 4.1.2 (Dihedral Group Configuration)**.** Consider the dihedral group D_4 of symmetries of a square, acting on the set $X = \{1, 2, 3, 4\}$. The group acts on $X \times X$ by $g(x, y) = (gx, gy)$ for any $g \in D_4$ and $x, y \in X$, and thus this action has three orbits:

$$\begin{aligned}
R_0 &= \{(1,1),(2,2),(3,3),(4,4)\} \\
R_1 &= \{(1,2),(2,3),(3,4),(4,1),(3,2),(2,1),(1,4),(4,3)\} \\
R_2 &= \{(1,3),(2,4),(3,1),(4,2)\},
\end{aligned}$$

and the corresponding adjacency matrices are given by

$$A_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_1 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

It is easy to see that $A_0 + A_1 + A_2 = J$, each matrix that is not A_0 has a zero diagonal and is symmetric, and

$$\begin{aligned}
A_1A_1 &= 2I + 0A_1 + 2A_2 \\
A_1A_2 &= A_2A_1 = 0I + 1A_1 + 0A_2 \\
A_2A_2 &= 1I + 0A_1 + 0A_2,
\end{aligned}$$

thus $\{A_0, A_1, A_2\}$ is a symmetric association scheme.

We will show that the previous example illustrates a general phenomenon: if G acts on X, we can always obtain a coherent configuration from a group by considering its *orbits*, that is, the orbits $R_0, ..., R_d$ of the induced action of G on $X \times X$ given by $g(x, y) = (gx, gy)$. This configuration is denoted by $\text{Inv}(G, X) := (X, \{R_i\}_{i=0}^d)$, and is called the *orbit configuration* of the group.

Theorem 4.1.3 (Orbit Configuration)**.** $\text{Inv}(G, X)$ is a coherent configuration. In particular, if $x \in X$, then the orbits of the stabilizer G_x of x in X are given by the sets $R_i(x)$, and $\text{Inv}(G, X)$ is an association scheme if and only if G acts transitively on X.

Proof. It is clear that the orbits $R_0, ..., R_d$ partition $X \times X$, and that if an orbit contains an element of the form (x, x), then all its elements are of the form (gx, gx) for $g \in G$, that is, this orbit will be a relation contained in $\{(x, x)|x \in X\}$, and if $R_i = G(x, y)$ is an orbit, then $R_{i'} = G(y, x)$ will be another orbit where $R_{i'} = R_i^T$. Thus, conditions (1), (2), and (3) of the definition of a configuration are satisfied. To show condition (4), note that for any $g \in G$ and $x, y \in X$, we have

$$\begin{aligned}(A_iA_j)_{g(x,y)} &= |R_i(gx) \cap R_{j'}(gy)| \\ &= |g(R_i(x)) \cap g(R_{j'}(y))| \\ &= |g(R_i(x) \cap R_{j'}(y))| \\ &= |R_i(x) \cap R_{j'}(y)| = (A_iA_j)_{xy},\end{aligned}$$

where the equalities follow from the fact that g is a permutation of the elements of X. Thus, A_iA_j is constant on the orbit of (x, y), and therefore it is a linear combination of the matrices A_l, proving that $\{A_0, ..., A_d\}$ is a coherent configuration. Naturally, the fibers of this scheme correspond to the orbits of G in X, and therefore G is transitive if and only if this configuration is homogeneous, i.e., if it is an association scheme. Finally, note that each set $R_i(x)$ is invariant under the action of G_x, so it is a disjoint union of orbits of G_x, but on the other hand, if $(x, y), (x, z) \in R_i$, then there exists $g \in G$ such that $g(x, y) = (x, z)$, so $g \in G_x$ and therefore $R_i(x)$ is contained in a single orbit of G_x, and it follows that $R_i(x)$ is an orbit of G_x. ■

The structure of the configuration $\text{Inv}(G, X)$ reflects some structural properties of the group G. In particular, we say that the action of G on X is *generously transitive* if for any $x, y \in X$, there exists $g \in G$ such that $gx = y$ and $gy = x$, that is, $g(x, y) = (y, x)$ and $g(y, x) = (x, y)$, and therefore $\text{Inv}(G, X)$ is symmetric if and only if G acts generously transitively. We say that G is 2-*transitive* if the action of G on the set $\{(x, y)|x, y \in X, x \neq y\}$ is transitive, that is, G is 2-transitive if and only if $\text{Inv}(G, X)$ has only two relations R_0, R_1. It can be proven that the symmetric group S_n acting on $\{1, ..., n\}$ is 2-transitive, so $\text{Inv}(S_n, \{1, ..., n\})$ will have only two relations whose adjacency matrices are I and $J - I$.

Now we will see another extremely important construction that relates groups and association schemes, called the *conjugacy class scheme*.

Theorem 4.1.4 (Conjugacy Class Scheme)**.** Let G be a group with conjugacy classes $C_0 = \{1\}, C_1, ..., C_d$, and define

$$R_i = \{(x, y)|x, y \in G, y^{-1}x \in C_i\}.$$

Then $(X, \{R_0, ..., R_d\})$ is a commutative association scheme.

Proof. First, we note that the group $H = G \times G = \{(x, y)|x, y \in G\}$ — viewed as a direct product of groups — acts on G as follows:

$$(x, y)g = xgy^{-1},$$

and consequently H also acts on the Cartesian product $G \times G$ via

$$(x, y)(g, h) = (xgy^{-1}, xhy^{-1}).$$

Our strategy will be to show that the orbits of this action are the sets R_i, and this, together with the previous result, will imply the result. To do this, first note that each R_i is invariant under the action of H, because if $(x, y) \in H$ and $(g, h) \in R_i$, we have

$$(xhy^{-1})^{-1}xgy^{-1} = yh^{-1}x^{-1}xgy^{-1} = (h^{-1}g)^{y^{-1}} \in C_i,$$

because $h^{-1}g \in C_i$ and C_i is a conjugacy class. On the other hand, H is transitive on R_i, because given $(g_1, h_1), (g_2, h_2) \in R_i$ where $h_1^{-1}g_1, h_2^{-1}g_2 \in C_i$, take $h \in G$ such that $(h_1^{-1}g_1)^h = h_2^{-1}g_2$, and define $x = h_1hh_2^{-1}, y = h$, and note that

$$(x, y)(g_1, h_1) = (h_2h^{-1}h_1^{-1}g_1h, h_2h^{-1}h_1^{-1}h_1h) = (g_2, h_2),$$

and it follows that each R_i is an orbit of H, and thus we indeed have an association scheme. To show that the scheme is commutative, consider a pair $g, h \in R_l$ where $h^{-1}g \in C_l$, and therefore $(h^{-1}g)^{h^{-1}} = gh^{-1} \in C_l$, implying that $(h^{-1}, g^{-1}) \in R_l$. Then, we consider the sets

$$F_1 = \{a \in G | (g, a) \in R_i, (a, h) \in R_j\} \quad \text{and} \quad F_2 = \{a \in G | (h^{-1}, a) \in R_j, (a, g^{-1}) \in R_i\},$$

where $|F_1| = p_{ij}^l, |F_2| = p_{ji}^l$, and then we consider the map $a \mapsto a^{-1}$ from F_1 to F_2, noting that $(g, a) \in R_i$ if and only if $(a^{-1}, g^{-1}) \in R_i$ and $(a, h) \in R_j$ if and only if $(h^{-1}, a^{-1}) \in R_j$, so the map is bijective, and therefore $p_{ij}^l = p_{ji}^l$, as desired. ■

The conjugacy class scheme is extremely important, and we will see in future sections that the eigenvalues of the matrices of this scheme are closely related to the irreducible representations of the group. Below we display the conjugacy class scheme of the group S_3.

Example 4.1.5. Let $S_3 = \{1, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}$, and note that this group has three conjugacy classes:

$$C_0 = \{1\}, C_1 = \{(1,2), (1,3), (2,3)\}, C_2 = \{(1,2,3), (1,3,2)\}.$$

We can explicitly compute the relations R_0, R_1, R_2, and from that, we obtain the matrices

$$A_0 = I, A_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and it follows that

$$\begin{aligned} A_1^2 &= 3I + 2A_2, \\ A_2^2 &= 2I + A_2, \\ A_1 A_2 &= A_2 A_1 = 2A_1. \end{aligned}$$

4.1.3 The Johnson and Hamming Schemes

Now, we will use the previous result to construct two extremely important association schemes: the Johnson scheme and the Hamming scheme. The Johnson scheme $\mathcal{J}(n, d)$ is constructed from a base set V of size n, and a constant d such that $n \geq 2d$. We define X as the set of all subsets of V of size d, and then we take

$$R_i = \{(x, y) | x, y \in X, |x \cap y| = d - i\},$$

that is, we group all pairs that have an intersection of size exactly $d - i$ in the relation R_i. The scheme is then given by $\mathcal{J}(n, d) = (X, \{R_i\}_{i=0}^d)$. Note that the group S_n acts on X as follows: if $x = \{x_1, ..., x_d\}$ and $\sigma \in S_n$, then

$$\sigma\{x_1, ..., x_d\} = \{\sigma x_1, ..., \sigma x_d\},$$

and with this, we will show that the orbits of this action are precisely the sets R_i, i.e., that $\mathcal{J}(n, d) = \text{Inv}(S_n, X)$. To do this, we will calculate the size of each R_i, and then use the Orbit-Stabilizer Theorem to conclude the result. But first, it is worth remembering that if G is a group acting on X, then a subset $S \subseteq X$ is an orbit if and only if S is G-invariant – that is, if $gs \in S$ for all $g \in G, s \in S$ –, and if G acts transitively on S.

Proposition 4.1.6. $\mathcal{J}(n, d)$ is a symmetric association scheme.

Proof. First, we show that the action of $G = S_n$ is generously transitive on X. Given $x, y \in X$ such that $(x, y) \in R_i$, we can write

$$x = \{x_1, ..., x_i, x_{i+1}, ..., x_d\} \quad \text{and} \quad y = \{y_1, ..., y_i, x_{i+1}, ..., x_d\},$$

that is, $\{x_1, ..., x_i\} \cap \{y_1, ..., y_i\} = \emptyset$, and then we define the permutation

$$\sigma = (x_1, y_1)(x_2, y_2)...(x_i, y_i),$$

that is, the product of transpositions that maps x_1 to y_1, x_2 to y_2, up to x_i to y_i, and fixes the other elements of X. It follows that $\sigma x = y$ and $\sigma y = x$, so the action is generously transitive, and thus each relation R_i is symmetric. We will now count the number of elements in each R_i to conclude that these are the orbits of G on X, and this, together with the previous result, will imply the desired result. If we fix an arbitrary element $x \in X$, there are

$$\binom{d}{i} \cdot \binom{n-d}{i}$$

elements $y \in X$ such that $(x, y) \in R_i$, that is, the cardinality of R_i is given by

$$|R_i| = \binom{n}{d} \cdot \binom{d}{i} \cdot \binom{n-d}{i}.$$

On the other hand, if we fix $(x, y) \in R_i$, there are $i! \cdot i! \cdot (d-i)! \cdot (n-d-i)!$ ways to construct a permutation that fixes (x, y), that is, this is the size of the stabilizer $G_{(x,y)}$, and then from the Orbit-Stabilizer Theorem we obtain that

$$|G(x, y)| = \frac{n!}{i! \cdot i! \cdot (d-i)! \cdot (n-d-i)!} = |R_i|,$$

and since $G(x, y) \subseteq R_i$ – that is, R_i is G-invariant –, it follows that R_i is an orbit, as we wanted. ■

The Hamming scheme $\mathcal{H}(d, q)$ is constructed from a set F with q elements – called the alphabet – and then we consider the set

$$X = F \times F \times \ldots \times F \quad \text{(d times)}$$

of tuples of size d with elements in F. We can think of elements of this set as strings of size d with elements from the alphabet F, e.g., if $q = 2$, then we are dealing with all binary strings of size d. If $x, y \in X$, we define

$$\partial(x, y) := |\{i | x_i \neq y_i\}|,$$

that is, $\partial(x, y)$ is a metric that calculates how many entries differ between x and y, and then we can consider the sets

$$R_i = \{(x, y) | x, y \in X, \partial(x, y) = i\}.$$

The Hamming scheme is then given by $\mathcal{H}(d, q) = (X, \{R_i\}_{i=0}^d)$. We can consider the group $S = (S_q)^d$ given by the direct product of d copies of S_q, and note that S acts on X via

$$(\sigma_1, ..., \sigma_d)(x_1, ..., x_d) = (\sigma_1 x_1, ..., \sigma_d x_d),$$

for any $\sigma = (\sigma_1, ..., \sigma_d) \in S, (x_1, ..., x_d) \in X$, that is, S acts on X by permuting each entry within the alphabet F. On the other hand, we can consider the group S_d acting on X by permuting the coordinates:

$$\sigma(x_1, ..., x_d) = (x_{\sigma^{-1}(1)}, ..., x_{\sigma^{-1}(d)}),$$

and we can note that the composition of these actions, that is, first permuting the entries according to S, and then permuting the coordinates, defines an action on X. If we consider S and S_d as subgroups of $\mathrm{Sym}(X)$, it follows that the first is normal but the second is not, and therefore we can consider the subgroup $G = SS_d$, which will be an internal semidirect product, that is,

$$G = S \rtimes S_d \subseteq \mathrm{Sym}(X).$$

We will show that $\mathcal{H}(d, q) = \mathrm{Inv}(G, X)$.

Proposition 4.1.7. $\mathcal{H}(d,q)$ is a symmetric association scheme.

Proof. Similarly to what was done for the Johnson scheme, we first show that the action of G on X is generously transitive. To do this, take elements

$$x = (x_1, ..., x_d) \quad \text{and} \quad y = (y_1, ..., y_d)$$

from X, and consider the permutation $\sigma \in S$ given by

$$\sigma = ((x_1, y_1), ..., (x_d, y_d)),$$

so that $\sigma x = y, \sigma y = x$, and therefore the action is indeed generously transitive. The set R_i is clearly G-invariant, so it remains to show that G acts transitively on R_i. Fix $(x,y),(w,z) \in R_i$, and assume that

$$\begin{aligned} x = (x_1, ..., x_i, x_{i+1}, ..., x_d) \quad &\text{and} \quad y = (y_1, ..., y_i, x_{i+1}, ..., x_d), \\ w = (w_1, ..., w_i, w_{i+1}, ..., w_d) \quad &\text{and} \quad z = (z_1, ..., z_i, w_{i+1}, ..., w_d), \end{aligned}$$

that is, we group the i distinct elements between x and y (resp. w and z) in the first i entries of the tuple, and the remaining common elements in the last $d-i$ entries. We can assume the elements are in this form since there always exists a permutation of the coordinates in S_d that maps any (x,y) to a tuple like the above. Then consider the permutation given by

$$\sigma = ((x_1, w_1), ..., (x_d, w_d)),$$

that is, σ maps each entry of x to the respective entry of w, so that $\sigma x = w$, and $\sigma y = (\sigma_1 y_1, ..., \sigma_i y_i, w_{i+1}, ...w_d)$, where $\sigma_i = (x_i, w_i)$. With this, we can take the permutation τ given by

$$\tau = ((\sigma_1 y_1, z_1), ..., (\sigma_i y_i, z_i), 1, ..., 1),$$

that is, τ maps the first i entries of σy to z, and fixes the others. But since the last $d-i$ entries of σy are equal to the last entries of w, which in turn are equal to those of z, it follows that $\tau\sigma y = z$. Now note that

$$\tau w = ((\sigma_1 y_1, z_1)w_1, ..., (\sigma_i y_i, z_i)w_i, w_{i+1}, ..., w_d),$$

so if $j \leq i$, the entries $(\sigma_j y_j, z_j)w_j$ will be different from w_j if and only if $w_j = z_j$ or $w_j = \sigma_j y_j$. The first case is false by assumption, and if the second is true, this would imply that

$$w_j = \sigma_j y_j = (x_j, w_j)y_j,$$

so $x_j = y_j$, which is also false by assumption, so $(\sigma_j y_j, z_j)w_j = w_j$, and thus $\tau w = w$. Therefore, we have $\tau\sigma(x,y) = (w,z)$, which means that G acts transitively on each R_i and therefore these are precisely the orbits of the action of G on X, as we wanted. ■

The construction of the group G used in the previous proposition illustrates a more general phenomenon that can be observed in groups, and which would be analogous to a kind of exponentiation. If G, H are groups, and if H acts on a finite set X of size d, we can consider the group G^d of tuples of size d of elements from G, and define an action of H on G by permuting the coordinates. From this, we can define the so-called *Wreath product* as

$$G \wr H := (G)^d \rtimes H.$$

The example of the Hamming scheme is an excellent motivation for this definition, as it provides a scenario where we can permute strings in two ways: (i) by permuting each entry within the alphabet, or (ii) by permuting the coordinates of the string.

4.2 Coherent Algebras

In the previous section, we saw some basic properties of configurations and schemes, but perhaps the most important thing about these objects is the fact that we can naturally associate an algebra to any configuration. If $\mathcal{C}$ is a configuration with adjacency matrices $A_0, \ldots, A_d$, we can consider the subspace

$$\mathcal{A} = \{\sum_{i=0}^{d} \alpha_i A_i \mid \alpha_i \in \mathbb{C}\}$$

generated by these matrices. This is, by construction, a subspace of dimension $d+1$ of $M_n(\mathbb{C})$, but note that by condition (4) of the definition of a configuration, the products A_iA_j can be expressed as a linear combination of the adjacency matrices, so $\mathcal{A}$ is a $\mathbb{C}$-algebra, with unit given by I and containing the matrix J. On the other hand, since the matrices A_i have entries in $\{0,1\}$ and $\sum_i A_i = J$, it follows that $A_i \circ A_j = 0$ if $i \neq j$, meaning that $\mathcal{A}$ is also an algebra with respect to the Schur product. Furthermore, since $A_i^T \in \{A_0, \ldots, A_d\}$, it follows that $\mathcal{A}$ is a $*$-algebra. The algebra $\mathcal{A}$ is called the *coherent algebra* associated with the configuration $\mathcal{C}$, and in the case of commutative schemes, this algebra is called the *Bose-Mesner algebra*. From these observations, we can note that the adjacency matrices of $\mathcal{C}$ are the Schur basis of the algebra in question, that is, the unique basis of primitive orthogonal idempotents with respect to the Schur product, and therefore, we can also refer to the adjacency matrices as the Schur basis of the configuration.

In general, we say that an algebra $\mathcal{A} \subseteq M_n(\mathbb{C})$ is coherent if it contains I, J, and is closed under the Schur product and the conjugate transpose map, that is, it is also a $*$-algebra and an algebra with respect to the Schur product. Using the results from Chapter 3, we can establish an equivalence between coherent algebras and configurations.

Proposition 4.2.1. If $\mathcal{A}$ is a coherent algebra, then its Schur basis is a coherent configuration.

Proof. By Lemma 3.3.1, we know that every algebra with respect to the Schur product has a unique Schur basis $A_0, \ldots, A_d$, and since $\mathcal{A}$ contains J, it follows that

$$\sum_{i=0}^{d} A_i = J.$$

Since $\mathcal{A}$ is an algebra, it follows that $A_iA_j \in \mathcal{A}$, and since these are 01 matrices, there exist non-negative integers p_{ij}^l such that

$$A_iA_j = \sum_{l=0}^{d} p_{ij}^l A_l.$$

Since $\mathcal{A}$ is closed under conjugate transpose, we have that $\{A_0^T, A_1^T, \ldots, A_d^T\}$ is a Schur basis for $\mathcal{A}$, so

$$\{A_0, \ldots, A_d\} = \{A_0^T, A_1^T, \ldots, A_d^T\},$$

meaning there exists some i' such that $A_i^T = A_{i'}$. Finally, given any A_i, since $I \in \mathcal{A}$, it follows that

$$A_i = A_i \circ I + A_i \circ (J - I),$$

where $A_i \circ I, A_i \circ (J-I) \in \mathcal{A}$ and $(A_i \circ I) \circ (A_i \circ (J-I)) = 0$, and since A_i is primitive, this implies that $A_i \circ I = 0$ or $A_i \circ (J-I) = 0$, meaning that if A_i has any diagonal entry, then A_i is a diagonal matrix, which concludes the proof. ∎

Therefore, every coherent algebra has a Schur basis that is a coherent configuration, and every coherent configuration is a Schur basis of a coherent algebra. From the results in the previous chapter, it follows that any coherent algebra is semisimple, meaning we can find its centrally primitive orthogonal idempotents $E_0, \ldots, E_m$ and decompose the algebra as

$$\mathcal{A} = \prod_{i=0}^{m} \mathcal{A}E_i,$$

and with this, it is possible to find an orthogonal basis that block-diagonalizes all the matrices in $\mathcal{A}$.

4.3 Commutative Schemes

Throughout this section, let $\mathcal{C} = (X, \{R_i\}_{i=0}^d)$ be a commutative association scheme where $|X| = n$, with Bose-Mesner algebra $\mathcal{A}$ and Schur basis $\{A_0 = I, A_1, \ldots, A_d\}$. Since $\mathcal{A}$ is commutative, we know from Theorem 3.3.2 that the primitive idempotents $E_0, \ldots, E_d$ will also form an orthogonal basis for the algebra $\mathcal{A}$, and by the construction of these idempotents, we have that $E_i^* = E_i$, and each E_i is an orthogonal projector onto a subspace of the matrices in the Schur basis. The fact that these two sets of matrices form orthogonal bases for $\mathcal{A}$ is perhaps the most useful property for applications, and now we will see how to derive algebraic relations between these bases and the parameters of $\mathcal{A}$.

The following basic properties about the parameters of a commutative association scheme can be demonstrated by simple counting arguments and algebraic manipulations, and therefore we leave the proof to the reader.

Proposition 4.3.1. For any indices $i, j, l, m \in \{0, \ldots, d\}$, the following hold for the parameters p_{ij}^l of a commutative association scheme:

(1) If $k_i := p_{ii'}^0$, then $k_0 = 1$, $k_i = k_{i'}, n = \sum_i k_i$, and $k_i \geq 0$;

(2) $p_{i0}^l = \delta_{li}$, $p_{0j}^l = \delta_{lj}$, $p_{ij}^0 = k_i \delta_{ij'}$, and $p_{ij}^l = p_{i'j'}^{l'}$;

(3) $k_i = \sum_{j=0}^d p_{ij}^l$ and $k_l p_{ij}^l = k_i p_{lj'}^i = k_j p_{i'l}^j$;

(4) $\sum_{t=0}^d p_{ij}^t p_{mt}^l = \sum_{t=0}^d p_{mi}^t p_{tj}^l$.

■

Naturally, in the case of symmetric schemes, $i' = i$ for any i, so the previous expressions simplify significantly. Now, we note that $(1/n)J$ will always be a primitive idempotent of $\mathcal{A}$, because if we assume without loss of generality that $JE_0 = \alpha_0 E_0$ with $\alpha_0 \neq 0$, then

$$\alpha_0 E_0 = JE_0 = \frac{1}{n} J^2 E_0 = \frac{1}{n} JE_0 J = \frac{\mathbb{1}^T E_0 \mathbb{1}}{n} J,$$

and using that $E_0^2 = E_0$, we can conclude that $E_0 = (1/n)J$. In the case of distance-regular graphs, this fact is quite natural: each matrix in the Schur basis induces a regular graph, so $\mathbb{1}$ is an eigenvector with a simple eigenvalue equal to the degree of that graph. We can also express the elements of the Schur basis in terms of the primitive idempotents, and vice versa. That is, there exist constants $P_{li}, Q_{li} \in \mathbb{C}$ such that

$$A_i = \sum_{l=0}^d P_{li} E_l \quad \text{and} \quad E_i = \frac{1}{n} \sum_{l=0}^d Q_{li} A_l,$$

and then, if we construct matrices $P, Q \in M_{d+1}(\mathbb{C})$ with entries $(P)_{ij} = P_{ij}, (Q)_{ij} = Q_{ij}$, we obtain that P is the change-of-basis matrix from the Schur basis to the idempotent basis, and $(1/n)Q$ is the change-of-basis matrix from the idempotent basis to the Schur basis. Thus,

$$PQ = nI = QP.$$

Since the idempotents are projectors onto subspaces, it follows that the i-th column of P contains the eigenvalues of A_i, that is, $A_i E_l = P_{li} E_l$, and the matrix P is therefore called the *first eigenmatrix of the scheme* (or also the first character table of the scheme). Similarly, it holds that $E_i \circ A_l = Q_{li} A_l$, and the matrix Q is called the *second eigenmatrix of the scheme* (or also the second character table of the scheme). It is important to observe that if we fix an idempotent E_l, the multiplicity of the eigenvalues P_{li} for the matrices A_i will be precisely $\operatorname{tr}(E_l) = m_l$, and these values are called the *multiplicities of the scheme*, and by definition, we have that $Q_{0l} = m_l$. With these observations, we prove the following statement:

Proposition 4.3.2. For any $l \in \{0, \ldots, d\}$, the following hold for the matrices P, Q of a commutative association scheme:

(1) $P_{l0} = Q_{l0} = 1$;

(2) $P_{0l} = k_l$ and $Q_{0l} = m_l$, where $k_l = p^0_{ll}$.

■

We recall that if A_i is one of the matrices in the Schur basis, there exists a unique index i' such that $A_i^T = A_{i'}$, and similarly, if E_i is a primitive idempotent, there exists a unique index $\hat{i}$ such that $E_i^T = E_{\hat{i}} = \overline{E_i}$. With this, we can prove the so-called orthogonality relations between the values of P and Q:

Theorem 4.3.3 (Orthogonality Relations)**.** The following relations hold for the matrices P, Q for any indices $i, j, l \in \{0, \ldots, d\}$:

(1) $P_{li} = \overline{P_{li'}}$;

(2) $Q_{li} = \overline{Q_{l\hat{i}}}$;

(3) $\overline{P_{li}} m_l = Q_{il} k_i$;

In particular, we have that

$$\sum_{l=0}^{d} \frac{P_{il}\overline{P_{jl}}}{k_l} = \begin{cases} \frac{n}{m_i}, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad \sum_{l=0}^{d} P_{li}\overline{P_{lj}} m_l = \begin{cases} nk_i, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. (1) It is enough to use that $A_{i'}E_l = P_{li'}E_l$ and apply the conjugate transpose to both sides.

(2) It is enough to use that $E_{\hat{i}} \circ A_l = Q_{l\hat{i}}A_l$ and apply the conjugate to both sides.

(3) First, note that $\operatorname{tr}(A_i E_l) = P_{li} m_l$, but on the other hand, we also have that $E_l = \sum_{j=0}^{d} Q_{jl} A_k$, so

$$A_i E_l = \frac{1}{n} \sum_{j=0}^{d} Q_{jl} A_i A_j = \frac{1}{n} \sum_{j=0}^{d} \sum_{t=0}^{d} Q_{jl} p^t_{ij} A_t,$$

and thus $\operatorname{tr}(A_i E_l) = \sum_{j=0}^{d} Q_{jl} p^0_{ij} = Q_{il} k_i$, as we wanted.
The last two relations are obtained by combining the fact that $PQ = QP = nI$ with item (3). ■

Now we note that since $\mathcal{A}$ is closed under the Schur product, we can find constants q^l_{ij} such that

$$E_i \circ E_j = \frac{1}{n} \sum_{l=0}^{d} q^l_{ij} E_l,$$

and these constants are the so-called *Krein parameters* of the scheme. We will soon show that these parameters are non-negative real numbers, but before that, we prove the following result:

Proposition 4.3.4. For any indices $i, j, l \in \{0, \ldots, d\}$, the following hold:

(1) $P_{li}P_{lj} = \sum_{t=0}^{d} p^t_{ij} P_{lt}$;

(2) $Q_{li}Q_{lj} = \sum_{t=0}^{d} q^t_{ij} Q_{lt}$;

(3) $P_{ji}Q_{lj} = \sum_{t=0}^{d} p^l_{it} Q_{tj}$;

(4) $p^l_{ij} = \frac{1}{nk_l} \operatorname{tr}(A_i A_j A_{l'})$;

(5) $q^l_{ij} = \frac{n}{m_l} \operatorname{tr}((E_i \circ E_j) E_l)$.

Proof. (1) It is enough to note that

$$P_{li}P_{lj}E_l = A_iA_jE_l = \sum_{t=0}^{d} p_{ij}^t A_tE_l = (\sum_{t=0}^{d} p_{ij}^t P_{lt})E_l.$$

(2) Similarly, we have

$$Q_{li}Q_{lj}A_l = E_i \circ E_j \circ A_l = \sum_{t=0}^{d} q_{ij}^t E_t \circ A_l = (\sum_{t=0}^{d} q_{ij}^t Q_{lt})A_l.$$

(3) From item (1), we have

$$P_{ji'}P_{jl} = \sum_{t=0}^{d} p_{i'l}^t P_{jt},$$

and from the previous proposition, we know that $P_{jl} = \frac{1}{m_j}k_l\overline{Q_{lj}}$ and $P_{jt} = \frac{1}{m_j}k_t\overline{Q_{tj}}$. Using these substitutions and taking the conjugate of both sides, we get

$$k_lP_{ji}Q_{lj} = \sum_{t=0}^{d} k_t p_{i'l}^t Q_{tj},$$

and then using that $k_t p_{i'l}^t = k_l p_{it}^l$, we obtain the desired result.

(4) Note that

$$A_iA_jA_{l'} = \sum_{t=0}^{d} p_{ij}^t A_tA_{l'} = \sum_{t=0}^{d}\sum_{m=0}^{d} p_{ij}^t p_{tl'}^m A_m,$$

so the trace of $A_iA_jA_{l'}$, which depends only on the coefficient of I in the sum on the right-hand side of the equation, will be

$$\operatorname{tr}(A_iA_jA_{l'}) = \sum_{t=0}^{d} np_{ij}^t p_{tl'}^0 = nk_l p_{ij}^l,$$

as required.

(5) Similarly, we have

$$(E_i \circ E_j)E_l = \frac{1}{n}\sum_{t=0}^{d} q_{ij}^t E_tE_l = \frac{q_{ij}^l}{n}E_l,$$

so

$$\operatorname{tr}((E_i \circ E_j)E_l) = \frac{q_{ij}^l m_l}{n},$$

which concludes the proof. ∎

It is easy to check from item (5) of the previous result that the Krein parameters are real numbers, and with a little more effort, it is also possible to check that they are always non-negative. However, since we do not need this fact in this work, we refer the reader to [Ter23, Cap. 4] for more details. To conclude this section, we will show a bound involving the multiplicities of a scheme:

Proposition 4.3.5. If $m_0, m_1, \ldots, m_d$ are the multiplicities of a commutative association scheme, then

$$\sum_{q_{ij}^l \neq 0} m_l \leq \begin{cases} m_i m_j, & \text{if } i \neq j, \\ \frac{m_i(m_i+1)}{2}, & \text{if } i = j. \end{cases}$$

Proof. Let $A, B \in M_n(\mathbb{C})$ be matrices with ranks m_A and m_B, respectively, and write

$$A = \sum_{i=1}^{m_A} v_i v_i^* \quad \text{and} \quad B = \sum_{i=1}^{m_B} x_i x_i^*,$$

that is, we write each matrix as a sum of linearly independent rank-1 matrices, and from this, we obtain that

$$A \circ B = \sum_{i,j} (v_i \circ x_j)(v_i \circ x_j)^*,$$

so $A \circ B$ is a sum of at most $m_A m_B$ linearly independent rank-1 matrices, which implies that $\mathrm{rk}(A \circ B) \leq \mathrm{rk}(A)\mathrm{rk}(B)$. If $A = B$, then we note that there are at most $\binom{m_A+1}{2}$ linearly independent rank-1 matrices, so $\mathrm{rk}(A \circ A) \leq \binom{\mathrm{rk}(A)+1}{2}$. Now, note that since the matrices E_i are orthogonal idempotents, it follows that

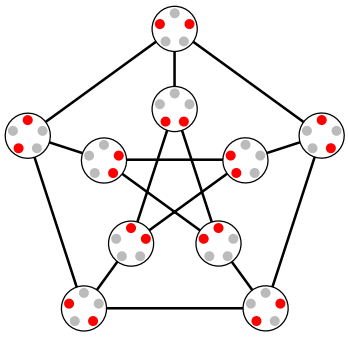
$$\mathrm{rk}(E_i \circ E_j) = \sum_{q_{ij}^l \neq 0} \mathrm{rk}(E_l) = \sum_{q_{ij}^l \neq 0} m_l,$$

and then, it is enough to combine this with the previous observations to conclude the result. ■

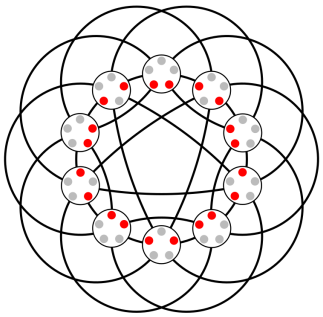
4.4 Distance-Regular Graphs

4.4.1 Definition and Basic Properties

The main example of association schemes are the so-called *distance-regular graphs*, and much of the theory about configurations and schemes arises from generalizations of properties exhibited by these graphs. To introduce the subject, we start with an example. In the previous section, we discussed the Johnson scheme $\mathcal{J}(n, d)$ formed by the set of d-subsets of a set with n elements, where the relation R_i is formed by pairs (x, y) whose intersection size is precisely $d - i$. This scheme is symmetric, meaning that we can view each matrix A_i associated with the relation R_i for $i > 0$ as a simple, undirected, and loopless graph. There are two graphs in this scheme that are particularly interesting: the graph A_d – where two sets x, y are related if, and only if, $|x \cap y| = 0$ – called the *Kneser graph* $K(n, d)$, and the graph A_1 – where two sets are related if, and only if, $|x \cap y| = d - 1$ – called the *Johnson graph* $J(n, d)$. The images in Fig.1 illustrate the case $\mathcal{J}(5, 2)$, where $J(5, 2)$ and $K(5, 2)$ are complementary.



(a) Petersen graph constructed as $K(5, 2)$.



(b) Johnson graph $J(5, 2)$.

Fig. 1: The two graphs obtained from the Johnson scheme $\mathcal{J}(5, 2)$.

On the other hand, we can note that $(x, y) \in R_i$ if, and only if, the distance between vertices x and y in $J(n, d)$ is precisely i, i.e., the length of the shortest path between these vertices in the Johnson graph is given by i. This means that we can also view the matrices $I, A_1, \ldots, A_d$ of the scheme $\mathcal{J}(n, d)$ as the distance matrices of the Johnson graph, that is, the matrices that relate vertices at distance i.

This example motivates the general definition of a distance-regular graph: if X is a connected k-regular graph with diameter d – that is, the size of the longest path in X is d – with adjacency matrix A, then we say that X is *distance-regular* if the matrices $\{I, A_1 = A, A_2, \ldots, A_d\}$ form a symmetric association scheme. Let

us now study the combinatorial implications of this definition. First, we note that the parameters $k_l = p^0_{ll}$ of the scheme are precisely the degrees of the graphs induced by the matrices A_i, i.e., the graph of A_i is k_i-regular, and we also note that since the distance in a graph is a metric, we can use the triangle inequality to obtain that if $p^l_{ij} \neq 0$, then $l \leq i+j$, $j \leq i+l$, and $i \leq j+l$, and in particular, $p^i_{j1} = 0$ if $j \notin \{i-1, i, i+1\}$. With this, we can define the main parameters for a distance-regular graph:

- $c_i = p^i_{(i-1)1}$, that is, given vertices u, v at distance i, c_i is the number of neighbors of v that are at distance $i-1$ from u, i.e., closer;
- $b_i = p^i_{(i+1)1}$, that is, b_i is the number of neighbors of v that are at distance $i+1$ from u, i.e., farther away;
- $a_i = k - b_i - c_i$, that is, a_i is the number of neighbors of v that are the same distance from u as v.

If we fix any vertex u and partition the other vertices of the graph according to their distance from u, we obtain the following diagram:

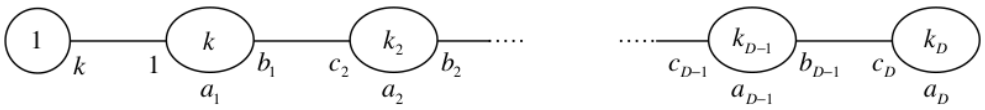


Fig. 2: Distance diagram [VDKT16, Figure 1].

It is always possible to perform this type of partition in any connected graph, but the fact that the graph is distance-regular shows us that the parameters of this diagram are the same regardless of which vertex u we fix. In fact, this is one possible definition of distance-regular graphs: if the parameters of the distance partition are the same regardless of the choice of the fixed vertex, then we say the graph is distance-regular. With this, we can count the number of edges between two components $i, i+1$ of the partition in question in two ways: first, we note that there are $k_i b_i$ edges from the component i to $i+1$, but on the other hand, there are $k_{i+1}c_{i+1}$ edges from the component $i+1$ to i, i.e., we obtain that

$$k_{i+1} = \frac{k_i b_i}{c_{i+1}},$$

and naturally, we have $n = 1 + k_1 + \cdots + k_d$. From now on, we will denote by X_i the relation given by the matrix A_i (and we will also treat X_i as a graph when necessary), where $(A_i)_{uv} = 1$ if the distance between u and v is precisely i, and similarly, we will denote by $X_i(u)$ the set of vertices adjacent to u in A_i. With this, we can show the following characterization of distance-regular graphs:

Theorem 4.4.1. Let X be a k-regular and connected graph with diameter d and adjacency matrix A. Then X is distance-regular if, and only if, there exist non-negative integers k_i, b_i, c_i such that $|X_i(v)| = k_i$ for any $v \in V(X)$, and for any $u \in X_i(v)$, it holds that

$$|X_1(u) \cap X_{i-1}(v)| = c_i,$$
$$|X_1(u) \cap X_{i+1}(v)| = b_i.$$

Proof. The forward direction is immediate from the previous considerations, so assume there exist integers k_i, b_i, c_i satisfying the stated properties, and define $a_i = k - b_i - c_i$. We now want to show that the matrices $I, A, A_2, \ldots, A_d$ form a symmetric association scheme, and for this, it is enough to check that condition (4) of the definition holds. It will be useful to define $A_{-1} = A_{d+1} = 0$ and leave c_{d+1}, b_{-1} free (their values will

not be relevant). With this, we note that

$$(AA_i)_{uv} = |X_1(u) \cap X_i(v)| = \begin{cases} b_{i-1}, & \text{if } u \in X_{i-1}(v), \\ a_i, & \text{if } u \in X_i(v), \\ c_{i+1}, & \text{if } u \in X_{i+1}(v), \end{cases}$$

that is, for any $i \in \{0, ..., d\}$, it holds that

$$AA_i = b_{i-1}A_{i-1} + a_iA_i + c_{i+1}A_{i+1}.$$

We can apply induction on i to show that A_i is always a polynomial in A of degree i, and also that $A^i \in \text{span}(\{I, A, A_2, ..., A_i\})$. In fact, the cases for $i = 0$ or $i = 1$ are immediate, and if $i = 2$, we have

$$A^2 = b_0I + a_1A + c_2A_2,$$

so A^2 is a linear combination of I, A, A_2, and since $c_2 \neq 0$ (because X is connected), we have that A_2 is a polynomial $\nu_2(A)$ of degree 2 in A, given by

$$c_2A_2 = c_2\nu_2(A) = A^2 - a_1A - b_0I.$$

If the result holds for i, then it follows clearly that

$$c_{i+1}A_{i+1} = c_{i+1}\nu_{i+1}(A) = A\nu_i(A) - a_i\nu_i(A) - b_{i-1}\nu_{i-1}(A),$$

so A_{i+1} is a polynomial of degree $i + 1$ in A. Now, let $A^i = \sum_{l=0}^{i} \alpha_l A_l$, where $\alpha_l \in \mathbb{C}$, then

$$A^{i+1} = \sum_{l=0}^{i} \alpha_l A_l A,$$

and using that A_lA is a linear combination of A_{l-1}, A_l, A_{l+1}, it follows that A^{i+1} is in $\text{span}(\{I, A, A_2, ..., A_{i+1}\})$, which concludes the induction. This shows that A_iA_j is a polynomial in A, and since

$$AA_d = b_{d-1}A_{d-1} + a_dA_d,$$

it follows that this polynomial has degree at most d, so $A_iA_j \in \text{span}(\{I, A, A_2, ..., A_d\})$, as required. ■

This result shows that a distance-regular graph X is completely determined by the parameters k_i, b_i, c_i, so we can define the *intersection array* of X as

$$\iota(X) := \{b_0, b_1, ..., b_{d-1}; c_1, ..., c_d\}.$$

In general, given an intersection array, it is not an easy task to determine whether there are distance-regular graphs with these parameters, and it is also difficult to determine if a distance-regular graph is the only one with a given array (up to isomorphism). The Petersen graph has an array given by $\{3, 2; 1, 1\}$, and it is possible to prove that it is determined by this array as a distance-regular graph, that is, any other distance-regular graph with this array is isomorphic to the Petersen graph. The smallest intersection array that corresponds to a pair of non-isomorphic graphs is $\{6, 3; 1, 2\}$, as both the Hamming graph $H(2, 4)$ and the Shrikhande graph have exactly this same array (for more details, we refer the reader to [VDKT16, Ch.2]).

4.4.2 Spectrum

Given a graph X with adjacency matrix A, we can consider the polynomial algebra $\mathbb{C}[A]$ in A — in this case, called the *adjacency algebra* of X — and note that it contains all the powers of A. Since A is the adjacency of a graph, the entries uv of A^r count the number of walks of length r between vertices u and v, that is, if d is

the diameter of X, then the matrices $I, A, A^2, ..., A^d$ form a linearly independent set in $\mathbb{C}[A]$. Therefore, the dimension of $\mathbb{C}[A]$ — which is precisely the number of distinct eigenvalues of A — is at least $d+1$, that is,

$$|\text{Dspec}(A)| \geq d+1,$$

where $\text{Dspec}(A)$ is the set of distinct eigenvalues of A. If X is distance-regular, we can consider the Bose-Mesner algebra $\mathcal{A}$ generated by the matrices $I, A, A_2, ..., A_d$, and then note that Theorem 4.4.1 tells us that $\mathcal{A} = \mathbb{C}[A]$, that is, the adjacency algebra of X is equal to the algebra generated by its distance matrices, since each A_i is a polynomial in A. We saw in Example 3.1.1 that the dimension of $\mathbb{C}[A]$ is precisely $|\text{Dspec}(A)|$, and with this we prove the following statement:

Proposition 4.4.2. If X is a distance-regular graph with diameter d and Bose-Mesner algebra $\mathcal{A} = \mathbb{C}[I, A, ..., A_d]$, then $\mathbb{C}[A] = \mathcal{A}$ and

$$|\text{Dspec}(A)| = d+1.$$

■

If we consider a distance-regular graph X with diameter d, we can define a tridiagonal matrix

$$L = \begin{pmatrix} a_0 & b_0 & 0 & 0 & \dots & 0 \\ c_1 & a_1 & b_1 & 0 & \dots & 0 \\ 0 & c_2 & a_2 & b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & c_{d-1} & a_{d-1} & b_{d-1} \\ 0 & 0 & \dots & 0 & c_d & a_d \end{pmatrix},$$

where $L \in M_{d+1}(\mathbb{R})$, known as the *intersection matrix*. We will now show that the eigenvalues of L coincide with the eigenvalues of A. The first important fact about L is that it is diagonalizable, because if we define the diagonal matrix $\Delta = \text{Diag}(k_0, ..., k_d)$, noting that

$$\frac{b_i\sqrt{k_i}}{\sqrt{k_{i+1}}} = \frac{c_{i+1}\sqrt{k_{i+1}}}{\sqrt{k_i}},$$

we conclude that $\Delta^{1/2} L \Delta^{-1/2}$ is symmetric. We observe that if $x = (x_0, \dots, x_d)$ is an eigenvector of L associated with the eigenvalue θ, then

$$\theta x_i = c_i x_{i-1} + a_i x_i + b_i x_{i+1}$$

for any $i \in \{0, ..., d\}$, where $c_0 = b_d = 0$, so we do not need to define x_{-1} and x_{d+1}. Additionally, we observe that $x_0 \neq 0$ (otherwise the previous recursion would give $x = 0$), so we can always assume that $x_0 = 1$, which in turn implies that $x_1 = \theta/k$ — this form of writing x is usually called the *standard sequence* with respect to θ. Now we can prove that L has exactly $d+1$ distinct eigenvalues and that $\text{Dspec}(A) = \text{Dspec}(L)$.

Proposition 4.4.3. If X is a distance-regular graph with diameter d, then

$$\text{Dspec}(L) = \text{Dspec}(A).$$

Proof. Let $x = (1, \theta/k, x_2, \dots, x_d)$ be the standard sequence with respect to θ, and fix a vertex v of X. Define an n-dimensional vector z by $z_u = x_{D(u,v)}$, for all $u \in X$, so that, if $\mathbf{a}_i$ denotes the v-th column of the i-th distance matrix A_i, we have

$$z = \sum_{i=0}^{d} x_i \mathbf{a}_i.$$

Now, we observe that

$$\begin{aligned} Az &= \sum_{i=0}^{d} x_i A\mathbf{a}_i \\ &= \sum_{i=0}^{d} x_i(b_{i-1}\mathbf{a}_{i-1} + a_i\mathbf{a}_i + c_{i+1}\mathbf{a}_{i+1}) \\ &= \sum_{i=0}^{d} (c_i x_{i-1} + a_i x_i + b_i x_{i+1})\mathbf{a}_i \\ &= \theta z, \end{aligned}$$

where the second equality follows from the identity

$$AA_i = b_{i-1}A_{i-1} + a_iA_i + c_{i+1}A_{i+1},$$

and the third equality follows by noting that each entry in $\mathbf{a}_j$ appears exactly three times in the sum when $i \in \{j-1, j, j+1\}$. This shows that $\mathrm{Dspec}(L) \subseteq \mathrm{Dspec}(A)$. Now, observe that

$$L - \theta I = \begin{pmatrix} -\theta & b_0 & 0 & 0 & \dots & 0 \\ c_1 & a_1 - \theta & b_1 & 0 & \dots & 0 \\ 0 & c_2 & a_2 - \theta & b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & c_{d-1} & a_{d-1} - \theta & b_{d-1} \\ 0 & 0 & \dots & 0 & c_d & a_d - \theta \end{pmatrix},$$

and since the c_i are all non-zero for $i \in \{1, ..., d\}$, it follows that $\mathrm{rk}(L - \theta I) \geq d$, so $\mathrm{null}(L - \theta I) \leq 1$, which implies that the eigenspace of each eigenvalue has dimension 1, so $|\mathrm{Dspec}(L)| = d + 1$, which completes the proof. ■

We can also calculate the multiplicities of each eigenvalue of A from the standard sequences of L.

Theorem 4.4.4 (Bigg's Formula)**.** Let X be a distance-regular graph with diameter d, and let $x = (1, \theta/k, \dots, x_d)$ be the standard sequence associated with an eigenvalue θ of L. Then, the multiplicity $m(\theta)$ of θ as an eigenvalue of A is given by:

$$m(\theta) = \frac{n}{\sum_{i=0}^{d} x_i^2 k_i}.$$

Proof. Let E be the projection onto the eigenspace of θ of A, and note that since X is distance-regular, $E \in \mathbb{C}[I, A, \dots, A_d]$, so we can write

$$\sum_{i=0}^{d} \alpha_i A_i.$$

We observe that the diagonal entries of E are constant and equal to α_0, so $m(\theta) = n\alpha_0$. We also observe that

$$\begin{aligned} \theta E = AE &= \sum_{i=0}^{d} \alpha_i AA_i \\ &= \sum_{i=0}^{d} \alpha_i(b_{i-1}A_{i-1} + a_iA_i + c_{i+1}A_{i+1}) \\ &= \sum_{i=0}^{d} (c_i\alpha_{i-1} + a_i\alpha_i + b_i\alpha_{i+1})A_i. \end{aligned}$$

On the other hand, $\theta E = \sum_{i=0}^{d} \theta\alpha_i A_i$, so

$$\theta\alpha_i = c_i\alpha_{i-1} + a_i\alpha_i + b_i\alpha_{i+1},$$

from which we conclude by induction on i that $\alpha_i = \alpha_0 x_i$. From the expression for E as a linear combination of the A_i's, we obtain

$$E^2 = \sum_{l=0}^{d} \sum_{i=0}^{d} \sum_{j=0}^{d} (\alpha_i \alpha_j p_{ij}^l) A_l,$$

and observing that $p_{ij}^0 = 0$ if $i \neq j$ and $p_{ii}^0 = k_i$, we conclude that the diagonal entries of E^2 are constant and equal to $\sum_{i=0}^{d} \alpha_i^2 k_i$. Since $E^2 = E$, and given that the diagonal entries of E are α_0, it follows that $\alpha_0 = \sum_{i=0}^{d} \alpha_i^2 k_i$. Using that $\alpha_i = \alpha_0 x_i$, we obtain

$$\alpha_0 = \frac{1}{\sum_{i=0}^{d} x_i^2 k_i}.$$

Thus, we conclude that

$$m(\theta) = \operatorname{tr}(E) = n\alpha_0 = \frac{n}{\sum_{i=0}^{d} x_i^2 k_i},$$

as desired. ■

The previous results show us that it is always possible to calculate the eigenvalues of A with their multiplicities from the matrix L. This is extremely practical, as it means that calculating the spectrum of A – which generally has a cubic cost in terms of the number of vertices – will have a cubic cost in terms of the diameter of X (which is generally much smaller than $|V(X)|$).

4.4.3 Imprimitivity

To conclude this section, we will discuss the concept of primitivity in distance-regular graphs. Given a distance-regular graph X, we say that it is *imprimitive* if any of its distance graphs X_i is disconnected (with $i > 1$), and primitive otherwise. Bipartite graphs with diameter at least 2 are always imprimitive, and the same holds for antipodal graphs – that is, graphs where X_d is a disjoint union of two or more cliques. The following result shows that any imprimitive distance-regular graph belongs to one of these families:

Theorem 4.4.5 (Smith's Theorem)**.** If X is an imprimitive distance-regular graph with degree $k > 2$, then it is bipartite or antipodal (or both).

Proof. Let X be as stated, with diameter d. We say that a triple of vertices $u, v, w \in V$ is of type (l, i, j) if $D(u, v) = l$, $D(u, w) = i$, $D(v, w) = j$, where D is the combinatorial distance in X. Note that if $j > 0$ and $p_{ii}^j > 0$, then any pair of adjacent vertices in X_j has at least one common neighbor in X_i, so any path in X_j generates a path in X_i. Therefore, if X_j is connected, then X_i must also be connected. If we choose i as the minimal index such that X_i is disconnected, it follows that if $j < i$, then $p_{ii}^j = 0$, i.e., there are no triples of type (j, i, i). Since X is connected by assumption, we have that $i > 1$.

Now, we consider three possible cases:

(i) If $i = d$, we show that X is antipodal. Note that $p_{dd}^j = 0$ for all $j < d$, so all triples of the type (j, d, d) are forbidden. This means that if $D(u, w) = D(u, v) = d$, then $D(w, v) = d$, i.e., X_d is a disjoint union of cliques, so X is antipodal;

(ii) If $2 = i < d$, we will show that X is bipartite. Indeed, we first consider a triangle wv_0v_1, and let $v_0v_1v_2v_3$ be a path of length 3 between vertices v_0, v_3 at distance 3. If we look at the triple w, v_0, v_2, then, since $D(w, v_0) = 1$ and $D(v_0, v_2) = 2$, it follows that $D(w, v_2) \in \{1, 2, 3\}$, but it cannot be 3, since wv_1v_2 is a path of length 2 from w to v_2, and it cannot be 2, since otherwise the triple would be of type $(1, 2, 2)$, which is forbidden, so it must be 1. Now, if we look at the triple w, v_1, v_3, similarly we have that $D(w, v_3) \in \{1, 2, 3\}$, but it cannot be 3, since wv_2v_3 is a path of length 2 from w to v_3, and it cannot be 1, since otherwise v_0wv_3 would be a path of length 2 between v_0 and v_3, so $D(w, v_3) = 2$ and wv_1v_3 is a forbidden triple of type $(1, 2, 2)$. This shows that there are no triangles in X. Now, let C be an odd cycle of length greater than 3, and note that all its vertices are in the same connected

component Δ with respect to X_2. If u, v are adjacent vertices in C and w is adjacent to u, then w cannot be adjacent to v, because there are no triangles, which means that w, v are adjacent in X_2, and thus w belongs to Δ. We can repeat this argument for any path connecting a vertex in X to C, concluding that all the vertices in the path must be in Δ, and since X is connected, this implies that X_2 is connected, which is a contradiction. Therefore, there are no odd cycles, and X is bipartite;

(iii) If $2 < i < d$, we will show that this leads to a contradiction. Consider a path of length d between vertices v_0, v_d at distance d, and note that since $k \geq 3$, we can find a vertex w adjacent to v_i that is distinct from v_{i-1}, v_i. Looking at the triple w, v_0, v_i, we obtain that $D(w, v_0) \in \{i-1, i, i+1\}$. If $D(w, v_0) = i$, then the triple w, v_i, v_0 is of a forbidden type $(1, i, i)$. If $D(w, v_0) = i+1$, then the triples w, v_i, v_1 and w, v_1, v_0 show that $D(w, v_1) = i$, so the triple w, v_{i+1}, v_1 is of a forbidden type (j, i, i) with $j \leq 2$. This implies that any neighbor of v_i distinct from v_{i+1} must be at distance $i-1$ from v_0, so $c_i = k-1$ and $b_i = 1$. Now, since the graph is distance-regular, it follows that v_{i+1} must also have c_i neighbors at distance $i-1$ from v_1, and since $k-1 \geq 2$, we can find a vertex z adjacent to v_{i+1} distinct from v_i, which is at distance $i-1$ from v_1. The triples z, v_0, v_1 and z, v_0, v_{i+1} show that $D(z, v_0) = i$, so $z \neq w$, and the triple z, v_i, v_0 is of a forbidden type (j, i, i), with $j \leq 2$.

The previous cases show that i is either d or 2, which in turn implies that X is either bipartite or antipodal, as we wanted. ■

The cycle with nine vertices C_9 is a counterexample to the previous theorem for $k = 2$, as it is imprimitive but neither bipartite nor antipodal, and the complete bipartite graphs $K_{d,d}$ are examples of imprimitive distance-regular graphs that are both bipartite and antipodal. As a final note, the previous theorem allows us to construct primitive graphs from imprimitive graphs. If X is an imprimitive bipartite graph with degree at least 3 and partitions V_1, V_2, then V_i is a connected component of X_2. The graphs induced by the components V_i in X_2 are called *halved graphs*, and are denoted by X^+, X^-. If X is antipodal, then we can obtain a graph X' with the vertex set given by the equivalence classes of $X_0 \cup X_d$, where two classes are adjacent if they contain vertices adjacent in X, and such a graph is called the *folded graph* of X. It can be shown that if X is distance-regular, then these graphs will also be distance-regular, and that after at most two steps of halving and/or folding, a primitive graph is obtained (we refer the reader to [BCN11, Ch.4] for more details).

4.4.4 Automorphisms

If X is a graph with n vertices, we can consider the group $G = \mathrm{Aut}(X)$ formed by the elements of $\mathrm{Sym}(n)$ that preserve adjacencies and non-adjacencies in X, i.e., by the permutations g such that $uv \in E(X)$ if and only if $g(u)g(v) \in E(X)$. This group is called the *automorphism group* of X. We will abuse notation and also identify G with its representation as a subgroup of $\mathrm{GL}(n, \mathbb{C})$ formed by permutation matrices (as in Example 2.5.1), and in this case, we can note that a permutation matrix P belongs to G if and only if $P^T AP = A$, where A is the adjacency matrix of X. The group G acts naturally on the vertices of X, and if this action is transitive, we say that the graph is *vertex-transitive*, and G also acts on the set of edges of X (where we understand that the edge ij is represented by the set $\{i, j\}$, so $ij = ji$), and if this action is transitive, we say that the graph is *edge-transitive*. The automorphism group also acts on the set of arcs of X, i.e., the set of ordered pairs of connected vertices (thus the arc (i, j) is different from the arc (j, i)), and if this action is transitive, we say that X is *arc-transitive*. Not every edge-transitive graph is arc-transitive, but every arc-transitive graph is necessarily both edge-transitive and vertex-transitive (for more information on graph automorphisms, we recommend [GR13, Ch.2] and [Big93, Chs.15-17]).

In this section, we will be interested in a particular case of graphs with high symmetry: the so-called *distance-transitive graphs*. If X is a connected graph with diameter d and automorphism group G, we say that it is distance-transitive if, for any vertices $u, v, w, z \in V(X)$ such that $D(u, v) = D(w, z)$, there exists $P \in G$ such that $P(u, v) = (P(u), P(v)) = (w, z)$, i.e., P maps u to w and v to z. The Johnson and Hamming graphs are examples of distance-transitive graphs, and it is easy to observe that from the above definition, every distance-transitive graph is also distance-regular. We will now see how to relate the orbital scheme with the Bose-Mesner algebra associated with a distance-transitive graph. First, we note that generally if

$P \in G$, then $P^T A_i P = A_i$ for any of the distance matrices A_i of a graph X, i.e., each A_i belongs to the centralizer algebra $C(G)$ of G in $M_n(\mathbb{C})$, and thus

$$\mathbb{C}[A] \subseteq \mathbb{C}[I, A, A_2, ..., A_d] \subseteq C(G),$$

for any connected graph of diameter d. We have seen that if X is distance-regular, then the first inclusion above is an equality, and now we will see that in the case of distance-transitivity, the chain of inclusions above is a chain of equalities. This will follow as a consequence of the following result:

Proposition 4.4.6. If G is a group of permutation matrices, then the coherent algebra $\mathcal{A}$ generated by the orbital configuration $\text{Inv}(G)$ is $C(G)$.

Proof. If A_i is one of the Schur basis matrices of $\mathcal{A}$ and if $P \in G$, then $P^T A_i P = A_i$, since the matrices A_i represent the orbitals of G. On the other hand, if $A \in C(G)$, it means that $P^T AP = A$ for any matrix $P \in G$, and this in turn implies that the 01 components of A will precisely be the orbitals of G, and therefore $P \in \mathcal{A}$. This shows us that $\mathcal{A} = C(G)$. ■

If G is the automorphism group of a distance-transitive graph, then by definition, we will have exactly $d+1$ orbitals, and this, together with the previous result, tells us that the dimension of $C(G)$ is precisely $d+1$, so $\mathbb{C}[A] = C(G)$. Finally, it is worth noting that if $\mathcal{A}$ is the coherent algebra generated by the conjugacy scheme of G, then $\mathcal{A} = Z(\mathbb{C}G)$, where $\mathbb{C}G$ is the algebra of the automorphism group viewed according to the representation of $\mathbb{C}G$ as permutation matrices (to see this, just note that the dimension of $\mathcal{A}$ is equal to the dimension of $Z(\mathbb{C}G)$ by Theorem 2.5.6).

4.5 Applications in Coding Theory

In this section, we will discuss some applications of association schemes to error-correcting code theory. This theory was developed independently of graph theory throughout the 20th century and has many practical applications of great importance. Also, throughout the 20th century, it became clear that this theory has many interesting intersections with other areas of mathematics, particularly with the theory of finite groups. The basic problem in this area is determining an optimal way to transmit a message through a medium that might corrupt it, either due to noise or external interference. This section contains only a brief introduction to the subject, and for further information, we refer the reader to [Coh74, Ch.11].

First, we begin with an alphabet F with q symbols, and then consider the set $X = F^n$ of words of length n over F, i.e., we start in the same way as the Hamming scheme seen in the last chapter. A code is simply a subset C of words in X. As a motivating example, suppose that Alice and Bob want to send a message through a channel, and for this, both agree beforehand on a code C. Alice then encodes her message and sends it to Bob. There are two basic tasks we can do in this context:

- **Error detection:** If Bob receives a message $x \in X$, he would like to determine whether any errors occurred in the transmission, and potentially the number of errors that occurred;
- **Error correction:** If Bob receives a message, he would like to detect and correct as many errors as possible.

To attempt to solve these tasks, we must introduce the concept of the *minimum distance* of the code C, which is defined as

$$\delta = \delta(C) := \min\{\partial(x, y) | x, y \in C, x \neq y\},$$

where ∂ is the Hamming distance. We will say that C is an (n, M, δ)-code if $C \subset F^n$, it has M elements, and the minimum distance is δ. The strategy we will adopt for detecting and correcting errors is as follows: if Bob receives a message x, we try to find a unique element $y \in C$ such that $\partial(x, y)$ is minimal. If this is possible, we can simply correct x to y, or return the distance from x to y as the number of errors that occurred. A good code is one that has a sufficiently large number of elements (to make encoding more efficient), but also has a large minimum distance (to make error detection and correction more efficient), and finding such codes is generally an extremely difficult task. With this, we can prove our first result:

Proposition 4.5.1. If C is an (n, M, δ)-code, then we can detect at most $\delta - 1$ errors and correct at most $\lfloor (\delta - 1)/2 \rfloor$ errors.

Proof. If $x \in C$ is a transmitted message and s errors occur resulting in the received message x', then $\delta(x, x') = s$. If $0 < s < \delta$, we will determine that x is the unique closest message to x' and detect the s errors, otherwise, we will not be able to detect the errors correctly. Now if $r \leq (\delta - 1)/2$ errors occur, we claim that x is the only element of C closest to x'. In fact, if there were another element $y \in C$ such that $\partial(x, x') = \partial(y, x') = r$, then

$$2r = \partial(x, x') + \partial(y, x') \geq \partial(x, y) \geq \delta \geq 2r + 1,$$

which is a contradiction, and thus we can correct the r errors. ■

Before continuing, we will discuss a simple example:

Example 4.5.2 (Parity Check)**.** Consider the binary alphabet $F = \{0, 1\}$, and suppose we want to send words of length n. We can then construct the following code: for each word $x \in F^n$, we append a 0 at the end if the number of 1's in x is even, and a 1 otherwise. In this way, we obtain a code C in F^{n+1} of size 2^n, and with minimum distance 2. By the previous result, this code can detect a single error but cannot correct any errors.

We will denote by $A_q(n, \delta)$ the largest size M of an (n, M, δ)-code, and if C is a code such that $M = A_q(n, \delta)$, we say that it is optimal, and if no larger code exists with the same minimum distance that contains C, we say it is maximal. We will define the closed ball of radius r centered at $x \in X$ as

$$B_r(x) := \{y \in X | \partial(x, y) \leq r\},$$

and we say that a set of balls is a *covering* of F^n if every point belongs to at least one ball, and it is a *packing* if every point belongs to at most one ball. We will now use these concepts to obtain bounds for $A_q(n, \delta)$. First, we will denote by

$$V_q(n, r) := |B_r(x)| = \sum_{i=0}^{r} \binom{n}{i} (q - 1)^i,$$

and thus $q^r = V_q(r, r) \leq V_q(n, r) \leq V_q(n, n) = q^n$. With this, we obtain the following result:

Theorem 4.5.3. If C is a maximal (n, M, δ)-code, then

$$\frac{q^n}{V_q(n, \delta - 1)} \leq |C| \leq \frac{q^n}{V_q(n, \lfloor (\delta - 1)/2 \rfloor)}.$$

Proof. To prove the result, we will construct a covering and a packing from a maximal code C as in the statement. First, we can consider the set $\{B_{\delta-1}(x) | x \in C\}$, and note that this will be a covering, because from the maximality of C it follows that for every $y \in F^n$ there exists some $x \in C$ such that $\partial(x, y) \leq \delta - 1$ (otherwise, we could enlarge the code). Therefore, we have that $|C| V_q(n, \delta - 1) \geq q^n$. For the upper bound, it is enough to consider the set $\{B_r(x) | x \in C\}$ with $r = \lfloor (\delta - 1)/2 \rfloor$, and note that if there exists some $y \in F^n$ such that $y \in B_r(x_1) \cap B_r(x_2)$, with $x_1, x_2 \in C$, then

$$2r \geq \partial(y, x_1) + \partial(y, x_2) \geq \partial(x_1, x_2) \geq \delta \geq 2r + 1,$$

which is a contradiction, so the set will be a packing, and from this it follows that $|C| V_q(n, \lfloor (\delta - 1)/2 \rfloor) \leq q^n$. ■

Since every optimal code is maximal, the previous result gives us bounds for $A_q(n, \delta)$. The previous results are classical in coding theory, and up to this point, they do not use the theory we developed in the last chapter. We will now see how to obtain a more general bound for $A_q(n, \delta)$ using association schemes, but for this, we will prove a general result about symmetric schemes, originally demonstrated by Delsarte [Del73].

Theorem 4.5.4. Let $(X, \{R_i\}_{i=0}^d)$ be a symmetric association scheme with second eigenmatrix Q, and let $C \subseteq X$ be any subset, and consider the distribution vector $a \in \mathbb{R}^{d+1}$ of C defined by

$$a_i = \frac{|(C \times C) \cap R_i|}{|C|},$$

i.e., the vector that counts the average degree in R_i of the subgraph induced by C.

Proof. First, note that if $\mathbb{1}_C$ is the indicator vector of the set C, then

$$a_i = \frac{\mathbb{1}_C^T A_i \mathbb{1}_C}{|C|},$$

and if $E_0, \ldots, E_d$ are the primitive idempotents of the scheme, then $E_i = (1/|X|) \sum_{l=0}^d Q_{li} A_l$, so

$$\begin{aligned} 0 &\leq ||\mathbb{1}_C^T E_i||^2 \\ &= (\mathbb{1}_C^T E_i)(\mathbb{1}_C^T E_i)^T \\ &= \mathbb{1}_C^T E_i \mathbb{1}_C \\ &= \frac{1}{|X|} \sum_{l=0}^d Q_{li} \mathbb{1}_C^T A_l \mathbb{1}_C \\ &= \frac{|C|}{|X|} (a^T Q)_i, \end{aligned}$$

and therefore $a^T Q \geq 0$. ■

Now consider the Hamming scheme $\mathcal{H}(n, q)$, which is symmetric, and let C be an (n, M, δ)-code. If we consider the distribution vector a of C, we have that $a \geq 0$, $a_0 = 1$, and $a_i = 0$ if $1 \leq i < \delta$, i.e., if we consider the linear program

$$\max\{\mathbb{1}^T y | y \in \mathbb{R}_+^{n+1}, y_0 = 1, y_i = 0 \text{ if } 1 \leq i < \delta, y^T Q \geq 0\},$$

we have that a is a feasible solution with objective value

$$\sum_{i=0}^n a_i = \frac{\mathbb{1}_C^T (\sum_{i=0}^n A_i) \mathbb{1}_C}{|C|} = |C|,$$

and with this we obtain the following result:

Theorem 4.5.5 (Delsarte Bound)**.** If C is an (n, M, δ)-code, then

$$|C| \leq \max\{\mathbb{1}^T y | y \in \mathbb{R}_+^{n+1}, y_0 = 1, y_i = 0 \text{ if } 1 \leq i < \delta, y^T Q \geq 0\},$$

where Q is the second eigenmatrix of the Hamming scheme $\mathcal{H}(n, q)$.

■

The previous result can be used to obtain bounds for the chromatic number and the size of a maximum coclique in strongly regular graphs (that is, distance-regular graphs of diameter 2, we refer to [BVM22] for more information), and it is considered one of the most important results in association scheme theory.

References

[Axl14] S. Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer International Publishing, 2014.

[Bai04] R. Bailey. *Association Schemes: Designed Experiments, Algebra, and Combinatorics*. Cambridge studies in advanced mathematics. Cambridge University Press, 2004.

[BCN11] A.E. Brouwer, A.M. Cohen, and A. Neumaier. *Distance-Regular Graphs*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2011.

[BGSV12] C. Bachoc, D. C. Gijswijt, A. Schrijver, and F. Vallentin. Invariant semidefinite programs. In M. F. Anjos and J. B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 219–269. Springer US, New York, NY, 2012.

[Big93] N. Biggs. *Algebraic Graph Theory*. Cambridge Mathematical Library. Cambridge University Press, 1993.

[BVM22] A.E. Brouwer and H. Van Maldeghem. *Strongly Regular Graphs*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2022.

[Cam99] P.J. Cameron. *Permutation Groups*. London Mathematical Society Student Texts. Cambridge University Press, 1999.

[Coh74] P.M. Cohn. *Algebra*. Number v. 2 in Algebra. Wiley, 1974.

[Coh12] P.M. Cohn. *Basic Algebra: Groups, Rings and Fields*. Springer London, 2012.

[CP23] G. Chen and I. Ponomarenko. Lectures on coherent configurations, October 2023.

[CR66] C.W. Curtis and I. Reiner. *Representation Theory of Finite Groups and Associative Algebras*. AMS Chelsea Publishing Series. Interscience Publishers, 1966.

[dCSCGR19] M. K. de Carli Silva, G. Coutinho, C. Godsil, and D. E. Roberson. Algebras, graphs and thetas. *Electronic Notes in Theoretical Computer Science*, 346:275–283, August 2019.

[Del73] P. Delsarte. *An Algebraic Approach to the Association Schemes of Coding Theory*. Philips journal of research / Supplement. N.V. Philips' Gloeilampenfabrieken, 1973.

[Die05] R. Diestel. *Graph Theory*. Electronic library of mathematics. Springer, 2005.

[Far12] D.R. Farenick. *Algebras of Linear Transformations*. Universitext. Springer New York, 2012.

[FD12] B. Farb and R.K. Dennis. *Noncommutative Algebra*. Graduate Texts in Mathematics. Springer New York, 2012.

[FIKW13] I.A. Faradzev, A.A. Ivanov, M. Klin, and A.J. Woldar. *Investigations in Algebraic Theory of Combinatorial Objects*. Mathematics and its Applications. Springer Netherlands, 2013.

[Gal21] J.A. Gallian. *Contemporary Abstract Algebra*. Textbooks in Mathematics. CRC Press, 2021.

[GM15] C. Godsil and K. Meagher. *Erdős–Ko–Rado Theorems: Algebraic Approaches*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2015.

[God93] C. Godsil. *Algebraic Combinatorics*. Chapman Hall/CRC Mathematics Series. Taylor & Francis, 1993.

[God10] C. Godsil. Association schemes, June 2010.

[GR13] C. Godsil and G.F. Royle. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer New York, 2013.

[Kar72] R. M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger, editors, *Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department*, pages 85–103. Springer US, Boston, MA, 1972.

[Lam13] T.Y. Lam. *A First Course in Noncommutative Rings*. Graduate Texts in Mathematics. Springer New York, 2013.

[Lan05] S. Lang. *Algebra*. Graduate Texts in Mathematics. Springer New York, 2005.

[LV16] M. Laurent and F. Vallentin. Semidefinite optimization, April 2016.

[Men23] L. Mendonça. Álgebra não-comutativa, Novembro 2023.

[Pas04] D.S. Passman. *A Course in Ring Theory*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2004.

[PdCSST23] N. B. Proença, M. K. de Carli Silva, C. M. Sato, and L. Tunçel. A primal-dual extension of the goemans–williamson algorithm for the weighted fractional cut-covering problem, 2023.

[Ter23] P. Terwilliger. Algebraic combinatorics: Association schemes, 2023.

[VDKT16] E. R. Van Dam, J. H. Koolen, and H. Tanaka. Distance-regular graphs. *The Electronic Journal of Combinatorics*, 1000, April 2016.

[VS21] A.C. Vieira and R.B. Santos. *PI-álgebras: uma introdução à PI-teoria*. 33º Colóquio Brasileiro de Matemática. IMPA, 2021.