CHAPTER III

Associative division algebras

3.1. Introduction.

After introducing sets, and a modification of the axioms, called mZFC, we document the standard rules for natural numbers and for a field, and then introduce zero algebras. We define magmas, groups, rings and ideals and determine some of their properties.

We use the hyperintricate representation for $2^n \times 2^n$ matrices to develop representations of the quaternions, and other algebras without complete division. We derive by hyperintricate methods that associative division algebras are of dimension ≤ 4. In characteristic zero these are the reals, complex numbers and quaternions.

On Wedderburn's little theorem, that any finite division ring is commutative, I give a proof using Lagrange's four squares theorem. To do this we introduce congruence arithmetic, and prove several preparatory theorems using it.

3.2. The description of sets.

A set is a collection of objects, like a set of socks in a drawer. The elements, or members, of this set are the individual socks, and the set is viewed abstractly, as all of the socks together. If S is a set and n is an element (or member) of S we write $n \in S$ (n belongs to S). If n does not belong to S we write $n \notin S$. A generalisation is to replace $\in$ by the mapping symbol $\rightarrow$ and to treat it as a type of function.

If a set S = {s, t, u, ... }, and all s, t, u $\in$ T, then if S is properly included in T we write $S \subset T$, and if there is the possibility that S = T then $S \subseteq T$. The intersection of two sets, S and T, is denoted by $S \cap T$ and satisfies if $s \in S$ and $s \in T$ then $s \in S \cap T$, whereas the union of two sets, S and T, denoted by $S \cup T$ satisfies if $s \in S$ or $s \in T$ then $s \in S \cup T$. The set complement of S in T, denoted by $\mathcal{C}S_T$ or T\S, satisfies the rule that if $s \notin S$ and $S \subset T$ then $s \in T$. The empty set, the set with no elements, {}, is denoted by $\varnothing$. The universal set $\mathbb{V}$ is the set for which every element belongs to it. This heresy is addressed in what follows.

The language of set theory has one binary relation symbol $\in$. A model of set theory provides a universe M and a binary relation E on M that interprets $\in$. The assumptions, or axioms, of a set may or may not include restricted Zermelo-Fraenkel set theory with the axiom of choice, ZFC:
Extension: For sets X and Y, X = Y whenever for all t, $t \in X$ if and only if $t \in Y$.
Empty set: There exists a set $\varnothing$ with $x \notin \varnothing$ for all x.
Pair: If x and y are members of a set, there exists a set T so that $t \in T$ whenever t = x or t = y.
Union: For every set X and Y, there exists a $u \in U$ (the union of X and Y) so that both $u \in X$ and $u \in Y$.
Power set: For all X, $T \subseteq X$ if and only if there is a set V with $T \subseteq V$.
Foundation: For all $X \neq \varnothing$ there exists a Y with the intersection $Y \cap X = \varnothing$.
Restricted comprehension: For every set B and every well-formed formula $\psi(x)$, there exists an $x \in X$ if and only if $\psi(x)$ holds and $x \in B$.

Axiom of infinity: There exists the set of natural numbers $\mathbb{N}_{\cup\varnothing}$ such that $\varnothing \in \mathbb{N}_{\cup\varnothing}$ and $x \in \mathbb{N}_{\cup\varnothing}$ implies $x \cup \{x\} \in \mathbb{N}_{\cup\varnothing}$.
Axiom of choice: Any of the usual formulations, for example: if $x \in X$ and $y \in Y \neq \varnothing$, then there exists a function f on x with $f(y) \in Y$ for all $y \in X$.

We adopt a modification of ZFC which we will call mZFC (*modified* ZFC) where we allow sets defined by properties. In mZFC, the *axiom of extended comprehension* holds: For every set B and every well-formed formula $\psi(x)$, either there exists an $x \in X$ if and only if $\psi(x)$ holds and $x \in B$, or the set B is the void set, $\odot$, for which $\psi(x)$ is identically false.

Thus if P is a logical property and relation, called a predicate, there exists a set $Y = \{x: Px\}$, where we allow $X = \odot$, the void set, which satisfies a false predicate. For example, if a set is defined as the set of all barbers in town, B, and $b \in B$ is defined by b cuts hair for everyone in town, and b also satisfies the property that b does not cut b's hair, then this is a contradiction and $B = \odot$. We do not take this as an antimony in the axiomatics of sets.

$\odot$ can be expressed indirectly as the complement of $\mathbb{V}$. Nonvoid sets obey a predicate which is somewhere true. For instance, applied to the set $\{X: X \notin X\}$ (Russell's paradox),

$$\odot = \{x: x \in \varnothing \text{ AND } x \notin \varnothing\}$$

satisfies the paradox. A universal set (or universe) is then the set for a true predicate, say

$$\mathbb{V} = \{x: x \in \mathbb{V} \text{ OR } x \notin \mathbb{V}\},$$

which is self-referential, in that $\mathbb{V}$ defines $\mathbb{V}$, or as a further example

$$\mathbb{V} = \{x: x = x\}.$$

The empty set satisfies $\varnothing \notin \varnothing$. If a set is void, it is empty: $\odot \subseteq \varnothing$, since a void set satisfies more conditions than an empty one. But $\varnothing$ satisfies the condition that no $x \in \varnothing$, and therefore $\odot \in \varnothing$ has a false $\psi(x)$, so $\odot = \varnothing$. Though we can form the set $\mathbb{V}'$ $\{x: \text{every } x \in \mathbb{V}'\}$, it is both the set complement of $\varnothing$ and $\odot$, so $\mathbb{V}' = \mathbb{V}$.

3.3. Natural numbers and integers.

We have seen the axioms for a set specify in the axiom of infinity that there exists a non-trivial set called the natural numbers. We will work with the set of natural numbers, $\mathbb{N}_{\neq 0} = \mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathbb{N}_{U0} = \mathbb{N} \cup \{0\}$ and the integers $\mathbb{Z}$ given by $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

Equality, =, satisfies the property that it is an *equivalence relation*, namely if m, n, p are members in a set then = is

reflexive: $m = m$ (1)
symmetric: if $m = n$ then $n = m$ (2)
transitive: if $m = n$ and $n = p$ then $m = p$. (3)

If it is not the case that $m = n$, then we write $m \neq n$.

If elements of a set S satisfy the property of being in an equivalence relation given by =, its application to elements which belong to S forms a *partition* of S:

if $m \in S$, then the intersection of all elements $n \in S$ which equal m, with the set of all $n' \in S$ which do not equal m is the empty set:

for every $m \in S$, $\{n; n = m \in S\} \cap \{n'; n' \neq m \in S\} = \varnothing$,

where also

for every $m \in S$, $\{n; n = m \in S\} \cup \{n'; n' \neq m \in S\} = S$.

The natural numbers satisfy the Peano axioms describing a recursive procedure to generate them.

$1 \in \mathbb{N}$ (4)

for every $n \in \mathbb{N}$, there exists an $S(n)$ interpreted as $(n + 1) \in \mathbb{N}$ (5)

there is no number $0 \in \mathbb{N}$ with $S(0) = 1$ (6)

for two numbers $m, n \in \mathbb{N}$, $S(n) = S(m)$ implies $n = m$. (7)

(induction) a subset of $\mathbb{N}$ containing 1 and $S(n)$ whenever $n \in \mathbb{N}$, is $\mathbb{N}$. (8)

We define $<$ by the property, if $m, n \in \mathbb{N}$, then $m < n$ if and only if there exists a $p \in \mathbb{N}$ so that

$m + p = n$.

[4Co80], [Sh91]. We expect that, for example, $S(S(S(n)))$ has only one preferred sequence of brackets, and we would normally define

$(SS)(S(n)) = S(S(S(n)))$, (9)

but if it is not the case that (9) holds, if $n = 1$ then additive associativity fails

$2 + (1 + 1) \neq (2 + 1) + 1$.

Since the parenthesis order of the right hand side of equation (9) is specified by the Peano axioms, but the left hand side is not, as an example we can specify

$(SS)(S(n)) = S'(S'(S(n)))$,

where $S'(n)$ is interpreted as $(n + 2)$, giving us a nonstandard nonassociative addition conforming to the Peano axioms. □

3.4. Fields and zero algebras.

The axioms for a *field* $\mathbb{F}, +, \times$, which we will denote simply by $\mathbb{F}$, satisfy for $a, b, c \in \mathbb{F}$, with $a \times b$ being written as ab

additive closure: $a + b \in \mathbb{F}$ (1)

associativity: $a + (b + c) = (a + b) + c$ (2)

abelian addition: $a + b = b + a$ (3)

existence of a zero: there exists a $0 \in \mathbb{F}$ satisfying

$a + 0 = a$ (4)

existence of negative elements: there exists a $(-a) \in \mathbb{F}$ with

$a + (-a) = 0$, (5)

which we write introducing subtraction as

$a - a = 0$

multiplicative closure: $ab \in \mathbb{F}$ (6)

associativity: $a(bc) = (ab)c$ (7)

commutativity: $ab = ba$ (8)

existence of a 1: there exists a $1 \in \mathbb{F}$ satisfying

$a1 = a$ (9)

existence of inverse elements: there exists an $a^{-1} \in \mathbb{F}$ for $a \neq 0$ with

$a(a^{-1}) = 1$, (10)

which we write introducing division as

$a/a = 1$

distributive law: $a(b + c) = (ab) + (ac)$. (11)

The motivation for introducing zero algebras is that multizeros which exist in them are consistent under division, whereas division by zero is inconsistent for a field. We now give the axioms for a zero algebra, which like the axioms for a set do not define uniquely its chosen elements, followed by a discussion of its properties and their models.

The axioms for a *zero algebra* $\mathbb{Y}$, under + and ×, which we will denote simply by $\mathbb{Y}$, satisfy conditions (1), (2), (3), (6), (7), (8), (9), (10) and (11) for a field with the axioms for zero excluded (so this means axiom (10) always holds in a zero algebra, since zero does not belong to it), and there exist multizeros $a0 \in \mathbb{Y}$ satisfying

ordered elements:	there exists an ordering on elements so that one of $a = b$, $a > b$ or $b > a$ holds, where $a \geq b$ and $b \geq c$ implies $a \geq c$	
additive cancellation:	$a = b$ if and only if $a + c = b + c$	(12)
negative pairing:	every a is paired with a (-a)	
multizeros:	if $a > (-a)$ an $a0 \in \mathbb{Y}$ exists with $a + (-a) = a0$,	(13)
which we write introducing subtraction in the case $a > (-a)$	$a - a = a0$	(14)
negative multiplication:	$(-a)b = -(ab)$	(15)
associativity:	$a(b0) = (ab)0$	(16)
existence of inverse elements for multizeros; division by multizeros satisfies:	$(a0)/(b0) = a/b$	(17)
existence of ultrainfinity:	$a/(b0) = (a/b)Ч$	(18)
ultrainfinity multiplication:	$(a0)(bЧ) = ab$	(19)
extension:	any property holding for, say, a above holds for the substitution a0 and aЧ.	(20)

For $\mathbb{Y}$ it may be convenient to represent a0 by a(0) when a is a specific decimal number.

The question arises of forming concrete models of zero algebras. We will look first at a standard model for them. Like complex numbers which are defined by their rules, being outside other number systems otherwise, we find zero algebras have no implementations in fields, but mappings from zero algebras to fields can be defined in cases where fields are consistent. We will also investigate whether nonstandard zero algebras exist.

The elements of a zero algebra denoted by ..., aЧЧ, aЧ, a, a0, a00, ... etc., may or may not be mapped to a field. There are two sorts of mappings from the operations on a zero algebra to the operations on a field: mappings for addition, and mappings for multiplication.

For the standard model of zero algebras, positive numbers in a zero algebra are mapped in sequence to the same positive numbers belonging to a field.

For additive operations in such a zero algebra using multizeros, the mapping to operations in a field is given by

$$a + (-a) = a0 \rightarrow a + (-a) = 0,$$

in the case when $a > (-a)$. This is the only such additive axiom.

For multiplicative operations in a standard zero algebra, the mapping to operations in a field are more varied:

$$a(b0) = (ab)0 \rightarrow a(b0) = (ab)0 = 0,$$

multiplication and division is closed under multizeros, and rules (12) and (15) operate in both a zero algebra and a field. $n0 = m0$ does not hold for a zero algebra when $n \neq m$, but does

hold in a field, where division by zero is inconsistent for a field, because 0.n = 0.m, so 0/0 = m/n for any m, n. This means

(n0)/(m0) = (n/m) → 0/0, which is not defined in a field.

Thus a zero algebra is consistent when its operations are constrained to operations consistent for a field, but it is also consistent for operations which are inconsistent for a field. □

We also have (-1)(-1) = 1 – 2(0), since

-1(1(0)) = -1(0)

(-1)(-1 + 1) = 1(0) – 2(0)

(-1)(-1) + (-1) = 1 – 1 – 2(0),

which proves the statement using the cancellation rule. Then compared with equation (13)

(-a) + (-a)(-1) = (-a) + a – 2a(0) = a(0) – 2a(0) = -a(0). (21)

Since axiom (10) for a field without a zero holds, we have

$(-1)(-1)^{-1} = 1$,

giving

$(-1)(-1)^{-1} – 2(0) = 1 – 2(0)$,

= (-1)(-1), (22)

and 𝕐 is consistent under subtraction, with a0 ≠ -a0, and we can say 0 ∉ 𝕐. □

There are degrees of freedom available to zero algebras, so a0 may be replaced by ay0, aЧ by aЧ/y where y ∈ 𝕐 with (ay0)/(by0) = a/b, etc. □

We can define a mapping between zero algebras, say in which for constant d

a′ = a, (-a′) = d + (-a),

where now under this transformation

a′ + (-a′) = d + a(0) = a′(0) (23)

for some number d ∈ 𝕐. This is a nonstandard model for a zero algebra. □

The complex numbers, ℂ, can be given a partial order as follows. Let $a + ib = re^{i\theta} \in \mathbb{C}$. Define an order on ℂ by

(1) c = a + ib > c′ = a′ + ib′ if the norm of c > the norm of c′. The norm is defined as the positive value of $\sqrt{a^2 + b^2}$ or alternatively and equivalently as the positive value of r.

(2) If the norms of c and c′ are equal, define c > c′ for those values for which a > a′.

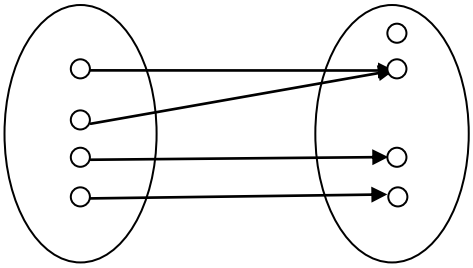
(3) If the methods (1) and (2) give the same result, define c > c′ for those values where b > b′.

It follows that otherwise the two complex numbers are equal.

Hence this order relation can be used to define a standard complex zero algebra. To define a complex zero algebra where the real part or the imaginary part are absent but not both, define these cases separately for the zero algebra. □

3.5. Magmas and groups.

A *function*, *mapping*, *map* or *transformation* f(x) is a set of pairs mapping from a set {x} of elements called the *domain* (or range) of the function to the set {f(x)} of elements called the *codomain* (or target, or *image*) of the function. The function can be represented by the arrow x → f(x) and is shown in the *cograph* diagram



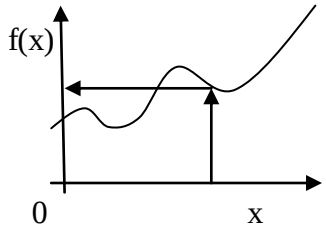
There are various types of order. A *partially ordered* set, for elements a, b and c in the set, has the relations of being

reflexive: $a \leq a$

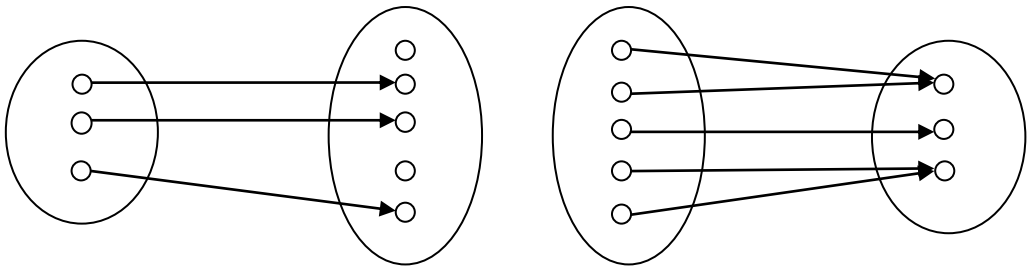
antisymmetric: if $a \leq b$ and $b \leq a$ then $a = b$

transitive: if $a \leq b$ and $b \leq c$ then $a \leq c$.

A function can also be represented by a set of *ordered pairs*, called the *Cartesian product* of the set {x} and {f(x)}, more generally of two sets S and T, and is denoted by $S \times T$, shown below in the *graph*



A function $S \rightarrow T$ is called *injective* (or *one-to-one*, or an injection) if $f(a) \neq f(b)$ for any two different elements a and b of the domain. It is called *surjective* (or *onto*) if f(S) = T. That is, it is surjective if for every element y in the codomain there is an x in the domain such that f(x) = y. The function f is called *bijective* if it is both injective and surjective.



injective function *surjective function*

A *dual* (or *opposite*) map reverses all arrows. If the original function is injective, then some of the elements of the opposite map may not have values in the codomain, so this is not a function on elements. Likewise, if the original function is surjective, the opposite map for an element in its domain may have not one but a set of elements corresponding to this element in its codomain, and again it is not a function on elements.

An *inverse function* (or *fiber*) of a bijective map $x \leftrightarrow f(x)$ is the map f^{-1}: $f(x) \leftrightarrow x$.

Sets of numbers may be combined under operations like '+' or '×' to form other numbers.

A *magma* is the most general structure combining sets and an operation. A magma is a set M with a single binary operation $M \times M \rightarrow M$, combining elements in pairs of the magma, with

each pair forming another element belonging to the magma. No other properties are specified in this definition.

A *group* is a magma with the following structure. The operation on the magma can be written either additively or multiplicatively, the two choices being equivalent within the group. The axioms for an *abelian group* (in which a + b = b + a) are given by axioms 3.4.(1) to (5) in the above fields and zero algebras section. A group is *nonabelian* or *noncommutative*, generally written multiplicatively, if some $ab \neq ba$, so 3.4.(6) to (10) but not (8) hold.

The *identity*, e, for a group if written additively is the element 0, or 1 written multiplicatively. The *inverse* of g written additively is -g, and g^{-1} if written multiplicatively.

In any group the integral *power* of an element a can be defined as the element

a^m = a.a. ... a (m terms).

Negative powers can be defined by

$a^m a^{-m} = 1$.

A group G is called *cyclic* if it contains an element the powers of which exhaust G. Cyclic groups are abelian. It is often understood that cyclic groups are finite. When this is not so, we will explicitly state that they are infinite.

A *permutation* is a bijection of a finite set to itself. A permutation which interchanges cyclically m objects of a set {1, 2, ..., m} forms an abelian group called a cycle of degree m. This permutation is obtained from a power by specifying that a^{m-1} is the m^{th} element and the cyclic permutation consists of multiplying by a. It can be represented by

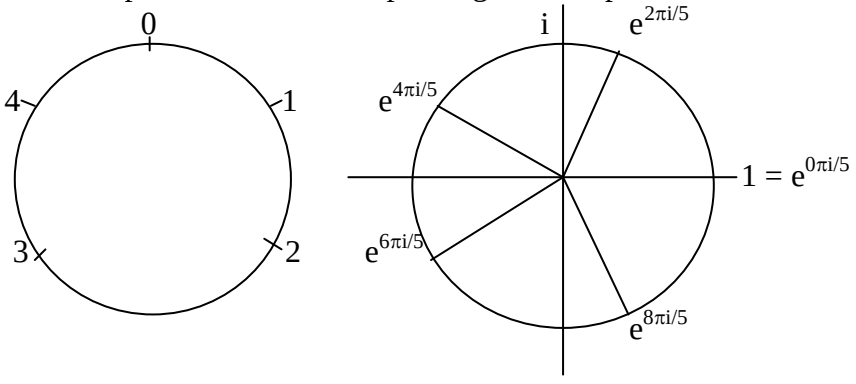
$$\begin{pmatrix} 1 & 2 & \cdots & m-1 & m \\ 2 & 3 & \cdots & m & 1 \end{pmatrix},$$

or in contracted notation by (1 2 ... m).

Another picture of a cyclic group is given by the 'clock' diagram x (mod p) and the Argand complex circle diagram, where there is a bijection for fixed r

$x \text{ (mod p)} \leftrightarrow e^{r + (2\pi i x/p)}$,

which can be pictured in the example diagrams for p = 5:



x (mod 5) complex (Argand) diagram with unit radius

A group derived from cyclic group generators which do not intersect, so the generators form a partition for the group, is also cyclic. For a finite cyclic group its number of elements, or *order*, which is the number of times it takes a generator to return to the identity permutation, is the least common multiple (l.c.m.) of the order of its cyclic components. An example of a cyclic permutation with the identity permutations present is

(1 2)(4 6 7)(3)(5)
which we can contract by removing the identity permutations to
(1 2)(4 6 7) = (4 6 7)(1 2).

The set of all permutations of m objects forms a group called the *symmetric group*, denoted by S_m. The name is derived from its origins in describing polynomial equations.

A noncommutative group can be generated by cycles which overlap somewhere. For example
$$\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}\begin{pmatrix}3&4\\4&3\end{pmatrix}=\begin{pmatrix}1&2&3&4\\2&4&1&3\end{pmatrix}\neq\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix}=\begin{pmatrix}3&4\\4&3\end{pmatrix}\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}.$$

The symmetric group can be described by matrices. For example $\begin{pmatrix}1&2&3&4&5\\2&4&1&3&5\end{pmatrix}$ can be represented by the matrix with one 1 in each row and column, and zeros elsewhere
$$\begin{bmatrix}0&1&0&0\\0&0&0&1\\1&0&0&0\\0&0&1&0\end{bmatrix},$$
in which, say, $2 \rightarrow 4$ is represented by a 1 in the second row and fourth column, with operations defined by matrix multiplication. As a further example, the cyclic group of order 4 is given by the matrices
$$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix},\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix}\text{ and }\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}.$$
All elements of S_4 can be obtained from the above by permuting rows or alternatively and equivalently by permuting columns.

A *subgroup* S of a group G is a group included in G. If $S \neq G$, S is a *proper subgroup*. The number of elements in the subgroup is called the order of the subgroup. The complement of S in G cannot form a subgroup, since 1 does not belong to it.

A *right coset or right residue class* of a subgroup S of G is the set of elements Sa, with $s \in S$ and $a \in G$. A *left coset* is the set aS, and when both coincide the set can be called a *coset*.

The *quotient group* G/S *of* G mod S for G a group, S a subgroup, is the family of left cosets.

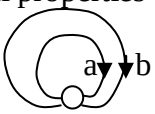
Lemma 3.5.1. *If* S *is finite, each right* (*or left*) *coset has as many elements as* S. *Two right* (*or left*) *cosets are either identical or have no common elements*.

Proof. The map $a \rightarrow sa$ is a bijection, since each sa is the image of one and only one a, and if $a \rightarrow sb$, with $b \neq a$ then $1 = a(a^{-1})$ maps to $sb(a^{-1}) = s(ba^{-1}) = s$, so b = a. Further, if there were any intersection, then sb = sa, which we have shown is impossible unless b = a. □

If G is finite, it can be partitioned into a finite number of right or left cosets where each coset contains the same number of elements, and the conclusion is
Theorem 3.5.2. (**Lagrange's subgroup theorem**). *The order of a finite subgroup* $S \subseteq G$ *divides the order of a finite group containing it*. □

A group can be thought of as not just as one operation acting on many elements, but also as one element with many operations.

The formal properties are the same, shown below.



The operations a, b and c (sometimes called *morphisms*) obey the rules of composition

$ab \in G$,

$a(bc) = (ab)c$,

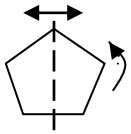
there is an identity operation 1 with

$1a = a = a1$

and there exists a reverse operation a^{-1} satisfying

$aa^{-1} = 1$. □

Considering a group as having one operation, this can also be split sometimes into a number of separate ones. An example is the dihedral group D_n with 2n elements, mentioned in chapter I, but shown here for the pentagon with n = 5. The two generators given by any standard rotation for n prime and a reflection about a vertical axis generate all elements of the group, which can be partitioned into five subsets each containing an unreflected and a reflected element, or two subsets, a reflected set with five rotated pentagons and a similar unreflected set.



A *homomorphism* h of a group G to a group G′ is a surjective map $ab = g \to h(g)$, such that $h(ab) = h(a)h(b)$.

Theorem 3.5.3. *Under a homomorphism the identity* e *of* G *maps to the identity* h(e) *of* G′, *and maps inverses* g^{-1} *to* $h(g)^{-1} = h(g^{-1})$.

Proof. The identity satisfies ee = e, so h(ee) = h(e) = h(e)h(e). The inverse satisfies $(g)(g^{-1}) = e$, so we get $h(g)h(g^{-1}) = h(e) = h(g)h(g)^{-1}$, and multiplying on the left by $h(g)^{-1}$ gives $h(g^{-1}) = h(g)^{-1}$. □

The set {k} is the *kernel* of a group homomorphism $h: G \to G'$, if it satisfies

$h(a)h(k) = h(a) = h(k)h(a)$,

in other words it is the identity of G′.

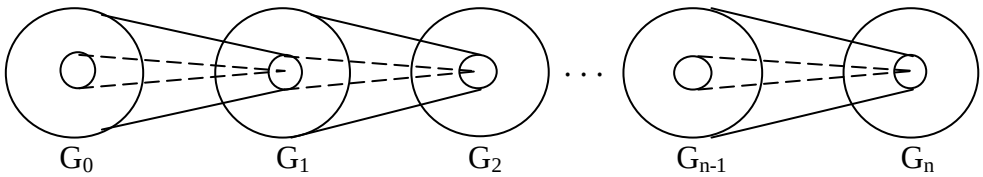
The following idea using kernels occurs in homology theory, discussed in *Number, space and logic* [Ad18]. A sequence of groups G_0, ..., G_n and homomorphisms f_1, ..., f_n is *exact* if the image (or the codomain) of each homomorphism is equal to the kernel of the next:

$$G_0 \xrightarrow{f_1} G_1 \xrightarrow{f_2} G_2 \xrightarrow{f_3} \ldots \xrightarrow{f_n} G_n,$$

that is

$im(f_k) = ker(f_{k+1})$.

We will show this again in the diagram



where the larger circle is the group, the smaller circle is its kernel, and the kernel maps to the identity. □

A bijective homomorphism is called an *isomorphism*. For example, the infinite cyclic group

$$G = \{1, a, a^{-1}, a^2, a^{-2}, ...\}$$

when written multiplicatively is isomorphic to the group $\mathbb{Z}$ of integers written additively.

An *automorphism* is an isomorphism of a group G to itself. It follows from theorem 3.5.3 that automorphisms form a group. □

A *conjugate* of an element x in a group G is an element $a^{-1}xa$.

Theorem 3.5.4. *For an element* a *of* G, *conjugation* T_a: $x \to a^{-1}xa$ *is an automorphism of* G.

Proof. $(a^{-1}xa)(a^{-1}ya) = a^{-1}(xy)a$. □

An automorphism of the form a^{-1} xa is called an *inner automorphism*, otherwise it is called an *outer automorphism*.

It follows from what we have said that inner automorphisms form a subgroup of all the automorphisms of a group G. □

A subgroup S of G is *normal* in G if and only if it is invariant under all inner automorphisms of G. For example, consider the symmetric group S_3 of all permutations of the set {1, 2, 3}. Then {(1, (1 2 3), (1 3 2)} is a normal subgroup of S_3, because we can verify the following statements.

(1 2){(1, (1 2 3), (1 3 2)} = {(1 2), (1 3), (2 3)} = {(1, (1 2 3), (1 3 2)}(1 2)
(1 3){(1, (1 2 3), (1 3 2)} = {(1 3), (2 3), (1 2)} = {(1, (1 2 3), (1 3 2)}(1 3)
(2 3){(1, (1 2 3), (1 3 2)} = {(2 3), (1 2), (1 3)} = {(1, (1 2 3), (1 3 2)}(2 3).

Theorem 3.5.5. *A subgroup* S *is normal if and only if all of its right cosets are left cosets*.

Proof. Suppose S is normal. Then $aSa^{-1} = (a^{-1})^{-1}S(a^{-1}) = S$. Thus Sa = aS. Conversely, applying lemma 3.5.1, if two cosets are equal so that Sa = bS, then a = b and S is normal. □

It should be carefully noted that the equation Sa = aS does not claim that every element of S commutes with a, only that the cosets Sa and aS are the same.

A group G is called *simple* if its only normal subgroups are the identity and G itself.

The classification of finite simple groups, completed in 2008, is a major milestone in the history of mathematics. Finite simple groups are classified as lying in one of 18 families, or otherwise as one of 26 exceptions:

The 18 families

- Z_p – a cyclic group of prime order.
- A_n – an alternating group for $n \geq 5$.
  The alternating groups may be thought of as groups of Lie type over the field with one element, which unites this family with the next, so all families of nonabelian finite simple groups may be considered to be of Lie type.

- One of 16 families of Lie type groups.
  The Tits group is usually considered of this form, although strictly speaking it is not of Lie type, but rather of index 2 in a Lie type group.

The 26 exceptions

- The sporadic groups, of which 20 are subgroups or subquotients of the Monster which are called the "Happy Family", while the remaining 6 are referred to as pariahs.

The famous theorem of Feit and Thompson states that every group of odd order is solvable. This means every finite simple group has even order unless it is cyclic of prime order.

3.6. Rings and ideals.

A *ring* has two operations, + and ×, satisfying for + the axioms of an additive group, for × in general the axioms of a noncommutative multiplicative group except that there may be no division, and the distributive laws connecting multiplication and addition, given in 3.4.(11).

The set of integers, $\mathbb{Z}$, which are the numbers ..., -3, -2, -1, 0, 1, 2, 3, ... can be formed into the *ring of integers*, which is the set $\mathbb{Z}$ with the usual operations of addition and multiplication. The letter $\mathbb{Z}$ originates from the German word *Zahl* for number.

The fractional numbers, like -1/3 or 22/7 form the *ring* $\mathbb{Q}$ *of rational numbers*, under the ordinary operations of addition and multiplication of fractions.

Similarly, the *real* or under a modified axiom system discussed in chapter VII the *Eudoxus numbers*, $\mathbb{U}$, with the ordinary operations of addition and multiplication form a ring $\mathbb{U}$.

The *complex numbers*, $\mathbb{C}$, numbers of the form a + bi, where a, b $\in \mathbb{U}$ and i = $\sqrt{-1}$ form a field, and therefore also a ring.

From the above rings as coefficients and variables we can form the set of polynomials together with the operations of addition and multiplication of polynomials, so from $\mathbb{Z}$ we can obtain the *polynomial ring* $\mathbb{Z}[x]$, for example

$(x^2 + x + 1) + (3x - 2) = x^2 + 4x - 1$

$(x^2 + x + 1)(3x - 2) = 3x^3 + x^2 - 2.$

Likewise, from the ring $\mathbb{Q}$ of rationals, we can form the polynomial ring $\mathbb{Q}[x]$, from $\mathbb{U}$ the polynomial ring $\mathbb{U}[x]$, and from the complex numbers $\mathbb{C}$ the polynomial ring $\mathbb{C}[x]$.

A further example of a ring is hyperintricate numbers (matrices), explained in chapters I and II. Unlike the previous examples, this is a noncommutative ring.

A *homomorphism of rings* for two rings A and A′ is a surjective map: A → A′: a → H(a) satisfying

$H(a + b) = H(a) + H(b),$ (1)

$H(ab) = H(a)H(b).$ (2)

By theorem 3.5.3 H satisfies

$H(0) = 0',$ (3)

$H(-a) = -(Ha),$ (4)

$H(a - b) = H(a) - H(b).$ □ (5)

A *two sided ideal* C in a ring A is a nonempty subset of A with the properties

$$\{c, d\} \in C \text{ implies } (c - d) \in C, \tag{6}$$

$$c \in C \text{ and } a \in A \text{ implies } ac \text{ and } ca \in C, \tag{7}$$

so an ideal of a ring A is a subset C of a that is itself a group under addition, and must satisfy an extra and strong condition with respect to multiplication – it must *absorb* multiplication from the ring A in the sense that if $c \in C$, then whenever we multiply c by an element $a \in A$, the product ac stays in C.

The set (2), also denoted by $2\mathbb{Z}$, of even integers is an ideal of the ring $\mathbb{Z}$ of integers.

Let A be a ring and let C be the set of all polynomials in the polynomial ring A[x] with zero *constant term* (coefficient a_0 missing), that is, if $f \in C$ then $f = a_n x^n + ... + a_2 x^2 + a_1 x$. Then the set C is an ideal of A[x]. To verify this, we first note that the difference between two polynomials with zero constant term is again a polynomial with zero constant term. Suppose $g = a_n x^n + ... + a_2 x^2 + a_1 x + a_0 \in A[x]$ and $h = b_m x^m + ... + b_2 x^2 + a_1 x \in C$, then we find

$$gh = a_n b_m x^{n+m} + ... + a_0 b_1 x \in C,$$

which means that C is an ideal of A[x].

Theorem 3.6.1. *Under any homomorphism* H *of a ring* A, *the set of all elements mapped onto zero is an ideal in* A.

Proof. Let C be the set of all elements in A with $H(c) = 0'$, with $0'$ the zero image in A'. If $H(c) = H(d) = 0'$, from (5)

$$H(c - d) = H(c) - H(d) = 0' + 0' = 0',$$

hence (6) holds, and (7) is proved by

$$H(ac) = (H(a))(H(c)) = H(a)0' = 0'. \square$$

An ideal of A need not be a subring of A, since it need not contain the unity 1.

Ideals can be used to create additional examples of rings. Just as we introduced cosets for groups, we can also introduce cosets for rings. Let C be an ideal of a ring A with a given element $a \in A$. A *coset* or *residue class* of C is a set of the form

$$a + C = \{a + c: c \in C\}.$$

The element a is called a *representative* of the coset a + C.

For example, let C = (3) be an ideal of $\mathbb{Z}$, the ideal also denoted by $3\mathbb{Z}$ of all multiples of 3. Then we can list the following cosets of C

$$0 + (3) = \{..., -6, -3, 0, 3, 6, ...\}$$
$$1 + (3) = \{..., -5, -2, 1, 4, 7, ...\}$$
$$2 + (3) = \{..., -4, -1, 2, 5, 8, ...\},$$

and using any other representative results in one of the three cosets listed above.

In general, we can form cosets of C, which are not subrings of A, so that

$$(a + C) + (b + C) = (a + b) + C,$$
$$(a + C)(b + C) = ab + C. \square$$

Theorem 3.6.2. *A homomorphic image of a ring* A *is determined up to isomorphism by the ideal of elements mapped to zero*.

Proof. We have to show that if H and K are homomorphisms of A onto rings A′ and A″, then H(a) = 0′ if and only if K(a) = 0″. Map a′ ↔ a″ whenever H(a) = a′ and K(a) = a″ for some a. This is a bijection, because firstly a′ and a″ are linked through at least one a, and secondly if a′ ↔ a″ and a′ ↔ b″, then H(a) = a′, K(a) = a″, H(b) = a′, K(b) = b″ for some a, b, so H(a – b) = a′ – a′ = 0′, which implies 0″ = K(a – b) = a″ – b″. Sums and products are also preserved by (6) and (7), for example a′ + b′ = H(a + b) = K(a + b) = a″ + b″. □

Theorem 3.6.3. *A division ring has no proper homomorphic images.*

Proof. Let C be an ideal with an element $c \neq 0$, then $1 = cc^{-1}$ is in the ideal and so is a = a1, so all elements of the ring are in the ideal. □

3.7. Representations of quaternions by hyperintricate numbers.

The quaternions are extensions of the complex numbers with 3 'imaginary' – or quaternionic – parts. So we can represent a quaternion by

$$a1 + bi + cj + dk$$

where

$$1^2 = 1,\ i^2 = j^2 = k^2 = -1,$$
$$1i = i = i1,\ 1j = j = j1,\ 1k = k = k1,$$
$$ij = k = -ji,\ jk = i = -kj,\ ki = j = -ik \qquad (1)$$

and the inverse is

$$(a1 – bi – cj – dk)/(a^2 + b^2 + c^2 + d^2). \qquad (2)$$

This (1, i, j, k) basis is representable by four hyperintricate numbers – in fact the previously given 1_1, α_i, i_1 and ϕ_i. An alternative representation, under swapping of layer levels, is 1_1, i_α, 1_i and i_ϕ. Some other representations are 1_{11}, $i_{\alpha\phi}$, $\alpha_{\phi i}$ and $\phi_{i\alpha}$ or 1_{1111}, $i_{11\alpha\phi}$, $\alpha_{ii\phi i}$ and $\phi_{iii\alpha}$. □

3.8. Non-existence of new associative division algebras.

Definition 3.8.1. An *associative division algebra* has a multiplicative identity $1 \neq 0$ where every nonzero element has a multiplicative inverse.

We prove that *the only associative division algebras are the reals, complex numbers and quaternions*. We will represent the basis elements of these associative division algebras by hyperintricate numbers.

In category theory, a basis of a vector space is an example of a universal arrow, which shows a vector has properties independent of basis. Nevertheless, we can also prove this is so in the case where the hyperintricate basis elements are transformable to the generalised case studied next.

These basis elements have square ±1, and any other representation can be reduced to a linear combination of these basis elements, for which the basis element squares are also ±1. In detail, any representation of the set {1, i, α, ϕ} under a change of basis which preserves squares maps each element to the set

$$\{1, \pm\sqrt{(\gamma_i^2 + \delta_i^2 + 1)}i + \gamma_i\alpha + \delta_i\phi, \pm\sqrt{(\gamma_\alpha^2 + \delta_\alpha^2 - 1)}i + \gamma_\alpha\alpha + \delta_\alpha\phi,$$
$$\pm\sqrt{(\gamma_\phi^2 + \delta_\phi^2 - 1)}i + \gamma_\phi\alpha + \delta_\phi\phi\}$$

with the coefficients $\gamma_i \neq \gamma_\alpha \neq \gamma_\phi$ etc. real. This extends to all layers.

It is not initially clear, for example, whether 1_{11}, $i_{\alpha 1}$, 1_{i1}, $i_{\phi 1}$, 1_{1i}, $i_{\alpha i}$, 1_{ii} and $i_{\phi i}$ can form a normed division algebra in which more than one square of a basis element is 1, e.g.1_{ii}.

For any hyperintricate basis element, the inverse is known. For a basis element A whose square is 1, the inverse $A^{-1} = A$. These A amount to all basis elements which have an even number (including zero) of i's in their hyperintricate representation. For any basis element, B, whose square is -1, the inverse $B^{-1} = -B$. The set of all B's is those basis elements which have an odd number of i's in their hyperintricate representation.

There are only two possibilities for basis elements, they either commute, AB = BA, or they anticommute, AB = -BA. This arises from the commutation or anticommutation of their layers.

Consider finding the inverse of $aA_1 + bA_2$, $A_1 \neq A_2$, where A_1^2 and $A_2^2 = 1$. This is then

$$(aA_1 - bA_2)/(a^2 - b^2) \qquad (1)$$

when A_1 and A_2 commute and

$$(aA_1 + bA_2)/(a^2 + b^2)$$

when A_1 and A_2 anticommute. If we use the fact that 1 is always present amongst such A's, then for some values of a and b, (1) holds, which implies that there exist a's and b's for which (1) includes the possibility of dividing by zero. The statement that we can do division is used in the definition of a division algebra (although we have to specifically exclude division by zero, as for a field), therefore there exists in such division algebras only one basis element with square 1, and this must be the real basis element. We will extend these considerations later.

To find the inverse of $a1 + bB_1$, where $B_1^2 = -1$, then this is

$$(a1 - bB_1)/(a^2 + b^2),$$

which introduces no further problems.

To find the inverse of $aB_1 + bB_2$, for $B_1^2 = -1$ and $B_2^2 = -1$, then this is the permitted

$$-(aB_1 + bB_2)/(a^2 + b^2),$$

when B_1 and B_2 anticommute, which is now the only possibility, since the alternative holds if and only if (1) can hold.

The above argument may be generalised for more B_r's, and it becomes necessary to stipulate that all B_1, B_2, ... B_n mutually anticommute.

We know there are solutions for B_1, B_2, B_3 given by basis elements for the quaternions. Now assume the existence of four such basis elements, B_1, B_2, B_3, B_4, all mutually anticommuting and distinct, so that $B_rB_s = -B_sB_r$. We will use associativity of these basis elements in computing from $B_1B_2B_3B_4$ its mirror reflection in two separate ways. So

$$\begin{aligned} B_1B_2B_3B_4 &= -B_1B_2B_4B_3 \\ &= B_1B_4B_2B_3 \\ &= -B_4B_1B_2B_3 \\ &= B_4B_1B_3B_2 \\ &= -B_4B_3B_1B_2 \\ &= B_4B_3B_2B_1. \end{aligned}$$

However

$$\begin{aligned} (B_1B_2)(B_3B_4) &= -(B_3B_4)(B_1B_2) \\ &= -(B_4B_3)(B_2B_1), \end{aligned}$$

a contradiction.

Thus the maximum number of dimensions for a standard associative division algebra is 4. □

3.9. Division with remainder. Congruence arithmetic.

To prove our version of Wedderburn's little theorem we will need to introduce congruence arithmetic, and then in the next section Lagrange's four squares theorem, where we prove a sequence of lemmas, elevated justly to theorems, to do this.

Strictly speaking, before proving the division with remainder theorem, we need to prove that *any number* n *greater than* 1 *is either prime or a product of primes*.

Proof. By the *method of induction* (also called *recursion*), if we wish to prove a statement for a positive whole number n, we can assume it has been proved for any number less than n. If n = 2, the theorem is proved. If n is composite, it can be represented as ab, where both a and b are greater than 1 and less than n, but we know by the induction method that a and b are either primes or the product of primes. □

The *fundamental theorem of arithmetic* states that *this factorisation is unique up to order of factors*.

Proof. If n is prime, the theorem is proved. Suppose n is composite and there are two factorisations

$$n = pqr ... = p'q'r' ...$$

Let these primes be ordered in increasing size. By the induction hypothesis, no prime p, q or r, ... can be the same as any of p′, q′, r′, ... otherwise we could divide n by it and get two representations of a smaller number, where we would continue the proof with this smaller number. Since n is composite it consists of at least two primes, so $n \geq p^2$ and $n \geq p'^2$ which implies $n > pp'$ with strict inequality, since p and p′ are unequal. Now form $n - pp'$. This has pp′ as a factor, so dividing pqr ... by p, p′ must be a factor of qr ..., a contradiction. □

Suppose we represent the integer parts of the rational numbers by notches along a line

```
|____|____|____|____|____|____|____|____|____|____|__
-1    0    1    2    3    4    5    6    7    8    9
```

then any rational number may be represented uniquely by an integer plus a rational number q equal to or greater than 0 and less than 1. This is an example of the theorem given by Euclid, in book 7 of the Elements.

The algorithm may be restated as:
Every positive whole number n *can be written uniquely in terms of a positive whole number* w *less than* n *multiplied by another natural number* $k > 0$ *with a unique remainder* $0 \leq t < w$:

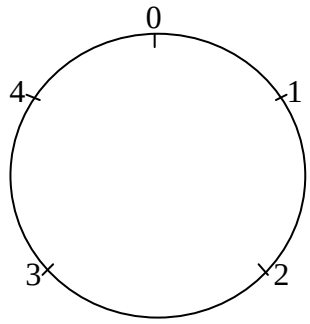
$$n = kw + t. \qquad (1)$$

Proof. If w divides n, then t = 0 and we are done. Otherwise assume (1). If n comes between kw and (k + 1)w, then (1) holds with $0 < t < w$. □

Using the theorem, we can develop an arithmetic for fixed w in which we only consider t above. This finite arithmetic is known as congruence, or clock arithmetic. The equation above is then written, in a notation due to Gauss

$$t = n \pmod{w},$$

which can be depicted in the example diagram for w = 5:



n (mod 5)

3.10. Theorems on sums of squares.

Theorem 3.10.1. (**Fermat's little theorem).** *For prime* p, *using the notation we have given,*

$$x^p - x = bp = 0 \pmod p \qquad (1)$$

for some unique b *dependent on* x.

Proof. We prove this by induction. For x = 0

$$0^p - 0 = 0p.$$

Assume (1) holds. Then for $x \to x + 1$, by the binomial theorem and the primality of p, so p does not divide any denominator

$$(x + 1)^p - (x + 1) = x^p - x + px^{p-1} + [p(p - 1)/2]x^{p-2} + ... + 1^p - 1 = bp + cp$$

for some unique c. □

Corollary 3.10.2. Fermat's little theorem may be rewritten from (1) as

$$x(x^{p-1} - 1) = x[(x^2)^{(p-1)/2} - 1]$$
$$= x(x^{(p-1)/2} - 1)(x^{(p-1)/2} + 1) = bp, \qquad (2)$$

so that if x is not zero or another multiple of p, for some unique r, s and t, squares in x, called *quadratic residues*, are of the form

$$x^{(p-1)/2} = 1 + rp \text{ if and only if } (x^2)^{(p-1)/2} = 1 + sp, \qquad (3)$$

otherwise non-squares are of the form

$$x^{(p-1)/2} = -1 + tp. \text{ □}$$

Corollary 3.10.3. Note that:

$$\text{if } (p - 1)/2 \text{ is even, then } x^{(p-1)/2} = (-x)^{(p-1)/2},$$
$$\text{but if } (p - 1)/2 \text{ is odd, then } x^{(p-1)/2} = -(-x)^{(p-1)/2}. \text{ □} \qquad (4)$$

Theorem 3.10.4. A theorem of Fermat states that *a prime* p *is expressible as the sum of two squares if and only if* $p \equiv 1 \pmod 4$ *or* p = 2. Generally, *a sum of two squares is a product of primes each of the form* $(4k + 1)^m$ *and* $(4k' + 3)^{2n}$ *possibly multiplied by a power of* 2.

Proof. We will at first assume the Fermat theorem and prove the general statement above.

We have $2 = 1^2 + 1^2$.

The square of any even number is congruent to 0 (mod 4) and of any odd number is congruent to 1 (mod 4). Therefore the sum, N, of two squares is congruent to (0 + 0), (0 + 1) or (1 + 1) (mod 4), and cannot be congruent to (4k + 3).

Since $(4k + 3)^{2n + 1} \neq (4r + 1)$ and $(4k + 3)^{2n} = (4r' + 1)$, sums of squares congruent to a product of primes must contain within the product at most terms of the form $(4k + 3)$ to an even power. But $(4k + 3)^{2n} = [(4k + 3)^n]^2 + 0^2$, and so is a sum of squares.

The Brahmagupta identity

$$(s^2 + bt^2)(u^2 + bv^2) = (su - btv)^2 + b(sv + tu)^2$$
$$= (su + btv)^2 + b(sv - tu)^2 \qquad (5)$$

with $b = 1$ implies inductively that any number N which is a product of primes 2, $(4k + 1)$ and $(4k' + 3)^{2n}$ can be represented as a sum of squares.

To prove the theorem of Fermat, we know $2 = 1^2 + 1^2$, and we need also to prove that prime $p = 4k + 1$ is expressible as the sum of two squares. We have seen from corollary 3.10.3 that for $p = 4k + 1$, that is, when $(p - 1)/2 = 2k$ is even, since 1 is a quadratic residue, so is -1, and this means the equation

$$x^2 + 1 = 0 \pmod p = mp$$

has a solution for $p = 4k + 1$ and some natural number m.

Suppose x lies between $\pm p/2$, ensured by subtracting from x a suitable multiple of p. Then

$$m = \frac{1}{p}(x^2 + 1) < \frac{1}{p}(\frac{1}{4}p^2 + 1) < p.$$

The idea of the proof is now as follows. For integers y and z let

$$mp = y^2 + z^2, \qquad (6)$$

with $m < p$. We will prove there is a natural number $m' < m$ with the same property. By repetition of the argument a finite number of times (this is known as Fermat's method of descent) we end up with an m' equal to one.

We will suppose u and v satisfy

$$-\frac{1}{2}m \leq u, v \leq \frac{1}{2}m$$

and

$$u = y \pmod m, v = z \pmod m, \qquad (7)$$

so that

$$u^2 + v^2 = y^2 + z^2 \pmod m,$$

and hence for an integer r

$$mr = u^2 + v^2. \qquad (8)$$

Note that $r \neq 0$, since u and v would then be zero, y and z would be multiples of m, so from (6) p would be a multiple of m, a contradiction. The size of r satisfies

$$r = \frac{1}{m}(u^2 + v^2) \leq \frac{1}{m}(\frac{1}{4}m^2 + \frac{1}{4}m^2) < m.$$

On multiplying equations (6) and (8) together, and using the Brahmagupta identity (5) with the variable $b = 1$

$$m^2rp = (y^2 + z^2)(u^2 + v^2) = (yu + zv)^2 + (yv - zu)^2, \qquad (9)$$

where by equation (7) the variables $(yu + zv)$ and $(yv - zu)$ are multiples of m, and so (9) is divisible by m^2, giving

$$rp = Y^2 + Z^2,$$

for some new variables Y and Z. Continuing this process, since we cannot finally end up with a value $r = 0$, or $r > 1$, we must terminate with $r = 1$, which proves the theorem. □

We will now extend these ideas further, proving Lagrange's four squares theorem.

Theorem 3.10.5. **Euler's four squares identity.** *The product of two numbers, each of which is a sum of four squares is itself a sum of four squares*.

Proof.

$$(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) =$$
$$(a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4)^2 + (a_1b_2 - a_2b_1 + a_3b_4 - a_4b_3)^2 +$$
$$(a_1b_3 - a_2b_4 - a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 - a_4b_1)^2. \square \qquad (10)$$

Theorem 3.10.6. **Lagrange's four squares theorem.** A*ny natural number is the sum of four squares*.

Proof. A number of versions, [Da82] chapter 5, [HW38] theorem 369, [La27] theorems 166-169 and [NZ60] paragraph 5.7, of Lagrange's proof can be found. The one below is mainly taken from Wikipedia, where the cases m even or odd do not use separate arguments.

It is sufficient to prove the theorem for every odd prime number $p = 4k + 3$. This follows immediately from Euler's four-square identity, and from the fact that the theorem is true for the numbers 1, 2 and that we have already proved the case for $p = 4k + 1$ in theorem 3.5.4.

Let a and b take integer values between 0 and $(p - 1)/2$ inclusive. The congruence

$$a^2 + b^2 + 1^2 + 0^2 = mp, \qquad (11)$$

with $0 < m < p$ has a solution, since by corollary 3.5.3, for $(p - 1)/2$ odd, that is, $p = 4k + 3$, then $-b^2 \neq 0$ is a non-residue (mod p) and since by definition a^2 is a quadratic residue, we want to find a residue R and a non-residue N satisfying from (11)

$$R + 1 = N.$$

Select the first non-residue in the sequence 1, 2, 3, ... Then the number before it is a residue, and we have satisfied (11).

Now let m be the smallest positive integer such that mp is the sum of four squares

$$mp = x_1^2 + x_2^2 + x_3^2 + x_4^2. \qquad (12)$$

We show by contradiction that m equals 1, which implies the theorem holds for p alone. Suppose this is not the case. We prove the existence of a positive integer r less than m, for which rp is also the sum of four squares, again in the spirit of the infinite descent method of Fermat.

For this purpose, we consider for each x_i the y_i which is in the same residue class modulo m and between $(-m + 1)/2$ and $m/2$ (included). It follows that $y_1^2 + y_2^2 + y_3^2 + y_4^2 = mr$, for some positive integer r less than m. Note as before that $r \neq 0$. The size of r satisfies

$$r = \frac{1}{m}(y_1^2 + y_2^2 + y_3^2 + y_4^2) \leq \frac{1}{m}(\frac{1}{4}m^2 + \frac{1}{4}m^2 + \frac{1}{4}m^2 + \frac{1}{4}m^2) = m.$$

This is not good enough as it stands, because we need to know that r is strictly less than m. If $r = m$, then y_1, y_2, y_3 and y_4 are all equal to $m/2$, so m must be even. But from (12) we would have mp divisible by m^2, so p would be divisible by m, which is impossible because p is prime and $1 < m < p$.

Finally, another appeal to Euler's four-square identity shows that $(mp)(mr) = z_1^2 + z_2^2 + z_3^2 + z_4^2$, where each z_i is divisible by m. Indeed, since each x_i is congruent to its corresponding y_i, z_1 is congruent modulo m to $y_1^2 + y_2^2 + y_3^2 + y_4^2 = mr$. For the same reason the other z_i are also divisible by m. It follows that, for $w_i = z_i/m$, $w_1^2 + w_2^2 + w_3^2 + w_4^2 = rp$, and this is in contradiction with the minimality of m. $\square$

Theorem 3.10.7. **Jacobi's theorem**. As further background, if $w^2 + x^2 + y^2 + z^2 = n$, then so is $w'^2 + x'^2 + y'^2 + z'^2$, where

$w' = ½(w + x + y + z)$
$x' = ½(w + x – y – z)$
$y' = ½(w – x + y – z)$
$z' = ½(w – x – y + z)$,

and the condition that w′, x′, y′ and z′ are whole numbers is that the number of even w, x, y and z is even.

Further, Jacobi found an exact formula for the total number of ways a given positive integer n can be represented in this way. Two representations are considered different if their terms are in different order or if the integer being squared (not just the square) is different.

The number of ways to represent n as the sum of four squares of positive, zero or negative integers is given by the symbol $r_4(n)$ and is eight times the sum of the divisors of n if n is odd, and 24 times the sum of the odd divisors of n if n is even, written

$r_4(n) = 8\sum_{\text{for m divides n}} m$

if n is odd and

$r_4(n) = 24\sum_{\text{for m divides n, m odd}} m$

if n is even. Equivalently, it is eight times the sum of all its divisors which are not divisible by 4, so that for a prime number p we have the explicit formula

$r_4(p) = 8(p + 1)$. □

3.11. Wedderburn's little theorem.

In this section we use a number of ideas developed in this chapter, including the properties of groups and rings, Lagrange's subgroup theorem and Lagrange's four squares theorem that every natural number is the sum of four squares.

Theorem 3.11.1. Wedderburn's little theorem states that *any finite division ring is commutative*. This invites some further remarks. It is a result of Feit and Thompson [FT63] that every finite group of odd order is solvable, or to put it another way, every nonabelian finite simple group has even order. Nonabelian finite groups may contain an element of order two (an involution). But the multiplicative part of a nonexistent finite nonabelian division ring containing such a group must then have two distinct basis elements with square 1, which we have seen separately is a contradiction, when these are not -1. When (1, -1) is a subgroup this is the abelian identity or kernel of a group called the *Schur multiplier* obtained from the nonabelian group, and we may apply the argument again to the new group if it remains nonabelian and find the contradiction without a distinct -1 being present. □

Proof. To obtain an independent proof of Wedderburn's little theorem, we first need to prove that every finite group can be represented by matrices (mod m). But every finite group can be represented by the permutation matrices of section 5, for which the (mod m) restriction is irrelevant. Thus multiplicatively a restriction to congruence arithmetic is bogus.

For a finite ring the additive part is abelian, and so can be mapped surjectively to (mod n) for some product $n = \prod_k n_k$ with cycles $(1, \ldots n_k)$ mapped bijectively to natural numbers n_k.

In general a division algebra is a subalgebra of the complex numbers, the quaternions or other nonassociative division algebras which may exist. Consider the associative case. We will see in chapter V and it is a result of Frobenius that the quaternions are the most general associative division algebra in infinitely countable arithmetic (a way of saying this is in *characteristic zero*), there being no such other algebras which are not subalgebras of these.

Let the order of the multiplicative part of a finite division ring be m. This is the order of the ring forgetting all additive operations. Let the order of the additive part of this ring be the n above, the order of the ring forgetting all multiplicative operations. Let the least common multiple (l.c.m.) of m and n be L. Then the order of the finite division ring is L. This follows from the (mod L) constraint being a necessary and sufficient condition for the distributive laws

$$a(b + c) = ab + ac \quad (1)$$

and

$$(f + g)h = fh + gh \quad (2)$$

to hold for the ring.

We prove that if

$$a = 0 \pmod m,$$
$$(b + c) = 0 \pmod n$$

and

$$m = pc$$
$$n = qc,$$

where c is the highest common factor of m and n, then

$$a(b + c) = 0 \pmod{pcq},$$

and pcq is the l.c.m. of m and n, since any divisor of pc and qc is a divisor of pcq, this holds maximally when the divisors are m and n respectively. Since (1) and (2) hold in characteristic zero, and we have shown that the left hand sides hold when say a = 0 (mod L) and (b + c) = 0 (mod L), by the Euclidean algorithm of section 9, the remainder $\leq$ L on division by a suitable multiple of L also satisfies (1) and (2).

If not all coefficients of the basis elements i, j, k of the quaternion

$$a1 + bi + cj + dk \pmod L$$

are zero, then the inverse is

$$a1 - bi - cj - dk/(a^2 + b^2 + c^2 + d^2) \pmod L,$$

and by Lagrange's four squares theorem, putting $L = (a^2 + b^2 + c^2 + d^2)$ means the inverse does not exist, since dividing by L is equivalent to dividing by zero. □

Corollary 3.11.2. This situation also obtains when L = m. □

Theorem 3.11.3. *Every nonzero Gaussian integer has an inverse* (mod n) *if and only if* n *is a prime* = 4k – 1.

If we have a torus of Gaussian integers a + bi, with coefficients (mod n) for 1 and (mod n) for i, the case for complex numbers is analogous. By theorem 3.10.4, since some numbers are not the sum of two squares, when this sum is absent (mod n) such non-zero inverses are always present for Gaussian integers (mod n).

The situation on one component, so that $a \equiv 0$ or $b \equiv 0$ but not both, is simpler. The condition now for the existence of all inverses $\neq 0$ is that n is prime, otherwise for g divides n there is

an h with gh = 0 (mod n), so g or h have no inverse. Thus overall a necessary and sufficient condition is that n = 3 (mod 4) is prime. □

3.12. Examples of possibly singular matrices.

If we were to allow equation 3.8.(1) to operate, this means that we can divide by $(a^2 - b^2)$. There exists the possibility that this is zero, but we could treat this situation on the same footing as dividing explicitly by zero, excluded as a number, but see paper II of [Ad14].

We will describe a nonstandard division algebra as one in which the number of singular occurrences, divided by the total number of occurrences, is a not well-ordered infinitesimal [Ad14], a number ε such that for any number $n \in N_{\neq 0}$ there does not exist an $m \in N$ with $\varepsilon m > n$. We will now incorporate these circumstances where $(a^2 - b^2) \neq 0$, which allows more than one timelike square, that is to say, we permit multiple basis element squares of 1.

If we write the intricate components possible for a basis as columns along a row and the layers as each row, then a quaternion basis is represented for instance by

1 i α ϕ
1 1 i i.

These basis elements are irreducible. To take an example

$(i + \alpha)_{(1 + i)} = i_1 + \alpha_1 + i_i + \alpha_i$,

where α_1 and i_i are not in this algebra.

Now consider the algebra which has some singularities as already described

1 i α ϕ
1 1 i i
1 i α ϕ
1 1 i i
_____
1 a b c,

where

$a^2 = b^2 = c^2 = 1$,
1a = a = a1, 1b = b = b1, 1c = c = c1,

and

ab = c = ba, bc = a = cb, ca = b = ac.

It is now possible to form a 16-dimensional associative algebra with a limited set of singularities, given by

1 i α ϕ	1 1 1	i i i	α α α	ϕ ϕ ϕ
1 1 i i	1 1 1	1 1 1	i i i	i i i
1 1 1 1	i α ϕ	i α ϕ	i α ϕ	i α ϕ
1 1 1 1	1 i i	1 i i	1 i i	1 i i
1 1 1 1	i α ϕ	i α ϕ	i α ϕ	i α ϕ
1 1 1 1	1 i i	1 i i	1 i i	1 i i,

and this process can be continued to produce 2^n-dimensional associative algebras with restricted division.

If we generalise the example for which we began the last section, we note that $1_{p,q,\ldots 1} + 1_{p,q,\ldots \alpha}$ is a matrix with a zero bottom row and therefore corresponds to a singular matrix. Similarly $1_{p,q,\ldots 1} + 1_{p,q,\ldots \phi}$ has two equal rows and is consequently also singular.

For the remainder of this section we will be considering $\{1_{p,q,\ldots 1}, \ldots\} \cup \{1_{p,q,\ldots i}, \ldots\}$, but we will see here too that a singular matrix can be derived with the trailing layer, in the example which follows by setting a = 1, g = 1 and all other coefficients zero.

To deal with the specific case considered next, first note that

$$P + Q1_{ii} + Ri_{\alpha i} + Si_{\phi i}$$

has inverse (other methods of finding hyperintricate inverses are given in chapter II)

$$(P - Q1_{ii} - Ri_{\alpha i} - Si_{\phi i})/(P^2 - Q^2 - R^2 - S^2). \quad (1)$$

Let us now investigate the properties of

$$\{1_{11}, 1_{i1}, i_{\alpha 1}, i_{\phi 1}, 1_{1i}, 1_{ii}, i_{\alpha i}, i_{\phi i}\}^{+,x} \quad (2)$$

under addition and multiplication. The above example is closed under multiplication. Does it form a multiplicative group?

We will write these 3-hyperintricate numbers as

$$a1_{11} + b1_{i1} + ci_{\alpha 1} + di_{\phi 1} + f1_{1i} + g1_{ii} + hi_{\alpha i} + ki_{\phi i}. \quad (3)$$

If we change (3) to an expression with inverses of basis elements substituted, we get

$$a1_{11} - b1_{i1} - ci_{\alpha 1} - di_{\phi 1} - f1_{1i} + g1_{ii} + hi_{\alpha i} + ki_{\phi i}. \quad (4)$$

Multiplying (3) by (4), a little manipulation gives the expression

$$[a^2 + b^2 + c^2 + d^2 + f^2 + g^2 + h^2 + k^2]$$
$$+ 2[(ag - fb)1_{ii} + (ah - fc)i_{\alpha i} + (ak - fd)i_{\phi i}], \quad (5)$$

which is precisely of the form (1).

Thus (3)×(4)×(1) = 1 provided

$$P = [a^2 + b^2 + c^2 + d^2 + f^2 + g^2 + h^2 + k^2],$$
$$Q = 2(ag - fb),$$
$$R = 2(ah - fi)$$

and

$$S = 2(ak - fd),$$

so that (3) does indeed have a multiplicative inverse. This is not a *normed* division algebra in the usual sense, since the denominator contains terms of degree 4. □

Further, we can continue such a process recursively, considering n-hyperintricate examples with trailing layer 1 or i, for instance derived from the above example. Let us look at this next. We will consider both the next stage up, and indicate how we can generalise in an induction procedure, and describe these in parallel. It is possible to be more formal, but then we can lose the thread of the idea.

The case corresponding to (2) is

$$\{1_{111}, 1_{i11}, i_{\alpha 11}, i_{\phi 11}, 1_{1i1}, 1_{ii1}, i_{\alpha i1}, i_{\phi i1}, 1_{11i}, 1_{i1i}, i_{\alpha 1i}, i_{\phi 1i}, 1_{1ii}, 1_{iii}, i_{\alpha ii}, i_{\phi ii}\}^{+,x}. \quad (6)$$

In an induction procedure, we assume a set of basis elements, and append as a trailing layer both 1 and i to those elements, thereby doubling the number of basis elements from its previous instance. By the induction procedure, there are 2^{n-1} elements to begin with, doubled to 2^n in the next stage.

Corresponding to (3), in the specific example we have chosen we consider the hyperintricate number

$$a1_{111} + b1_{i11} + ci_{\alpha 11} + di_{\phi 11} + f1_{1i1} + g1_{ii1} + hi_{\alpha i1} + ki_{\phi i1}$$
$$+ a'1_{11i} + b'1_{i1i} + c'i_{\alpha 1i} + d'i_{\phi 1i} + f'1_{1ii} + g'1_{iii} + h'i_{\alpha ii} + k'i_{\phi ii}, \quad (7)$$

whereas corresponding to (4), we generate the hyperintricate

$$a1_{111} - b1_{i11} - ci_{\alpha 11} - di_{\phi 11} - fl_{1i1} + g1_{ii1} + hi_{\alpha i1} + ki_{\phi i1}$$
$$- a'1_{11i} + b'1_{i1i} + c'i_{\alpha 1i} + d'i_{\phi 1i} + f'1_{1ii} - g'1_{iii} - h'i_{\alpha ii} - k'i_{\phi ii}. \qquad (8)$$

In the general situation we will have coefficients in lower case of hyperintricate numbers with a trailing 1 layer, minus, in the format corresponding to (8), hyperintricates in primed lower case coefficients each with basis element with a trailing i layer.

Multiplying (7) and (8) together gives

$$[a^2 + b^2 + c^2 + d^2 + f^2 + g^2 + h^2 + k^2]$$
$$+ [a'^2 + b'^2 + c'^2 + d'^2 + f'^2 + g'^2 + h'^2 + k'^2]$$
$$+ 2[(ag - fb - a'g' + f'b')1_{ii1}$$
$$+ (ah - fc - a'h' + f'c')i_{\alpha i1} + (ak - fd - a'k' + f'd')i_{\phi i1}]$$
$$+ 2[(ab' - a'b - fg' + gf')1_{i1i}$$
$$+ (ac' - a'c - hf' + fh')i_{\alpha i1} + (ad' - a'd - kf' + fk')i_{\phi 1i}]. \qquad (9)$$

In general there is a set of positive squares of coefficients both primed and unprimed, followed by twice a number of coefficients times basis elements with an even number of i's.

The inverse of

$$P + Q1_{ii1} + Ri_{\alpha i1} + Si_{\phi i1} + T1_{i1i} + Ui_{\alpha 1i} + Vi_{\phi 1i}$$

is

$$(P - Q1_{ii1} - Ri_{\alpha i1} - Si_{\phi i1} - T1_{i1i} - Ui_{\alpha 1i} - Vi_{\phi 1i})$$
$$/(P^2 - Q^2 - R^2 - S^2 - T^2 - U^2 - V^2), \qquad (10)$$

this being a generalisation of (1), involving in its typical characteristic basis elements an even number of i's, so that once again, in its general form, the inverse of (9) can be obtained, as is also derived using the result for a general hyperintricate inverse in chapter II, section 16. □

3.13. Exercises.

(A) Show that a quaternion represented in section 3.7 by $a1_1 + bi_1 + c\alpha_i + d\phi_i$ has determinant $[a^2 + b^2 + c^2 + d^2]^2$, where this is the square of the denominator in the equation for the inverse 3.7.(2).

(B) The next problem is about zero algebras, which have been described in section 3.4.

For an associative algebra, if b has a multiplicative inverse, show that

$$ab = cb$$

if and only if a = c.