Mathematics 111 — Algebra 2002–2003

Colm Ó Dúnlaing

April 9, 2003

1 Sets and maps

Contents of this section.

(1.1) Cantor's intuitive definition of a set. A *set* is a collection of objects considered as a whole.

(1.2) Various sets of numbers. The *natural numbers* $\mathbf{N}$ are the nonnegative integers:

$$\mathbf{N} = \{0, 1, 2, \ldots\}$$

The *integers* $\mathbf{Z}$ are the whole numbers, positive and negative

$$\mathbf{Z} = \{\ldots - 3, -2, -1, 0, 1, 2, \ldots\}$$

The *rationals* $\mathbf{Q}$ consist of all fractions

$$\mathbf{Q} = \{\frac{p}{q} \colon p, q \in \mathbf{Z} \text{ and } q \neq 0\}$$

The *reals* $\mathbf{R}$ consist of every number which is the limit of a Cauchy-convergent sequence of rationals (this description must suffice).

The *complex numbers* $\mathbf{C}$ are

$$\mathbf{C} = \{x + iy \colon x, y \in \mathbf{R}\}$$

where $i^2 = -1$.

The *quaternions* or *hypercomplex numbers* $\mathbf{H}$ are

$$\mathbf{H} = \{w + ix + jy + kz :: w, x, y, z \in \mathbf{R}\}$$

where $i^2 = j^2 = k^2 = ijk = -1$ (the Broombridge formula).

(1.3) Definition *A* map *with* domain A *and* codomain B *is a rule or procedure which associates with each $x \in A$ a unique element of B.*

The words 'function, mapping, transformation' are synonyms for 'map.'

(1.4) Notation *The notation $f\colon A \to B$ means that A and B are sets and f is a map with domain A and codomain B.*

Given $x \in A$, $f(x)$ denotes the unique element of B associated to x by f.

Alternatively, one can write $f\colon x \mapsto y$ to mean that y is the unique element of B associated with x by f.

(1.5) The identity map ι_X. If X is any set, there is a well-defined *identity map*

$$\iota_X\colon X \to X; \quad x \mapsto x$$

for every $x \in X$.

Other examples. Squaring: $\mathbf{R} \to \mathbf{R}; x \mapsto x^2$ is the map which squares each real number x. Parity

$$\mathbf{Z} \to \mathbf{Z}; \quad x \mapsto \begin{cases} 0 & \text{if } x \text{ is even} \\ 1 & \text{if } x \text{ is odd} \end{cases}$$

(1.6) Definition *Union:* $A \cup B = \{x\colon x \in A \text{ or } x \in B\}$
Intersection: $A \cap B = \{x\colon x \in A \text{ and } x \in B\}$
Difference: $A \backslash B = \{x\colon x \in A \text{ and } x \notin B\}$
Symmetric difference: $A \triangle B = \{x\colon (x \in A \text{ and } x \notin B) \text{ or } (x \in B \text{ and } x \notin A)$

(1.7) Definition *The* cartesian product $A \times B$ *of two sets A and B is the set of ordered pairs*

$$A \times B = \quad \{(x, y)\colon x \in A, y \in B\}$$

$A^2 = A \times A$*; generally,* $A^n = A \times A \times \cdots \times A$ *(cartesian product of n copies of A).*

Explanation. We use 'ordered pair' in an intuitive sense, just as in coordinate geometry, where points in the plane correspond to ordered pairs of real numbers.

The main idea about an ordered pair (a, b) is that one can identify its first element a and its second element b, and two ordered pairs are equal if and only if their first elements are equal and their second elements are equal.

The word 'cartesian' is related to 'cartesian coordinates.' $\mathbf{R} \times \mathbf{R}$ is the set of cartesian coordinates of points in the plane.

The multiplication sign reflects the fact that if A and B are finite then the cardinality $|A \times B|$ of the cartesian product is the product $|A| \times |B|$ of their cardinalities.

As mentioned in class, $\{\{a\}, \{a, b\}\}$ will work as a definition of (a, b).

	0	1	2
0	(0,0)	(0,1)	(0,2)
1	(1,0)	(1,1)	(1,2)
2	(2,0)	(2,1)	(2,2)

	0	1	2
0	0	1	2
1	1	2	3
2	2	3	4

Figure 1: $A \times A$ where $A = \{0, 1, 2\}$, and part of the addition table

Examples of cartesian products. If $A = \{1, 2\}$ and $B = \{a, b, c\}$, then

$$A \times B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

Notice $|A| = 2$, $|B| = 3$, and $|A \times B| = 6$.

If $A = \{x \in \mathbf{R} \colon 1 \leq x \leq 3\}$ and $B = \{x \in \mathbf{R} \colon 4 \leq x \leq 9\}$, then $A \times B$ corresponds to a rectangular planar region of area 10.

More examples of maps. Operations like addition and subtraction can be viewed as certain kinds of map. For example, addition of natural numbers can be considered as a map from $\mathbf{N} \times \mathbf{N}$ to $\mathbf{N}$, where $(x, y) \mapsto x + y$. Elements (x, y) of the cartesian product $\mathbf{N} \times \mathbf{N}$ can be viewed as places in the addition table. In Figure 1, the cartesian product of $\{0, 1, 2\}$ with itself is shown, together with the addition table for these particular numbers.

Similarly, multiplication, $\log_e$, exponentiation, etcetera, can be defined as maps with suitable domains and codomains.

Restrictions and extensions of maps.*This topic has been deleted.*

Example. *Deleted.*

(1.8) Equality of maps. Two maps $f \colon A \to B$ and $g \colon C \to D$ are equal if and only if $A = C$, $B = D$, and for all $x \in A$, $f(x) = g(x)$. They may be given be by completely different formulae.

Example. *The Gamma function example has been deleted.* Suppose $A = C = IN$, $b = D = \mathbf{Z}$, $f(x) = (-1)^x$, and $g(x) = 1 - 2x + 4 * (x \div 2)$. Then $f = g$.

(1.9) Range (or image) of a map. Given $f \colon A \to B$, the *range of* f is the set $\{f(x) \colon x \in A\}$. It is a subset of B but not necessarily all of B.

Given $X \subseteq A$, we write $f(X)$ for $\{f(x) \colon x \in X\}$ so $f(A) = \text{range}(f)$.

(1.10) Compatible maps and composition. Let $f \colon A \to B$ and $g \colon C \to D$ be two maps. If $\text{range}(f) \subseteq \text{domain}(g)$ (i.e., C), then we say g is *compatible with* f and we define the *composite map* $g \circ f$, *g following f*, as follows:

$$g \circ f \colon A \to D; \quad x \mapsto g(f(x)).$$

When g is compatible with f we may simply say that the composite map $g \circ f$ is *defined.*

Example. If $f(x) = x + 1$ and $g(x) = x^2$, with $A = B = C = D = \mathbf{R}$, then

$$g \circ f(x) = (x + 1)^2 \quad \text{and} \quad f \circ g(x) = x^2 + 1.$$

Composition of maps is not commutative. This example shows that $g \circ f$ and $f \circ g$ need not be equal, even when both are defined. That is, composition of maps is not a commutative operation.

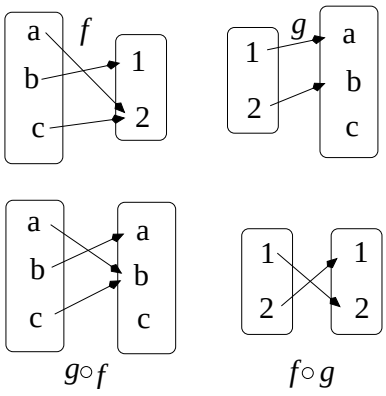


Figure 2: f, g, $g \circ f$, $f \circ g$.

However, it is associative. For example, suppose we are considering three maps f, g, h, all with domain $\mathbf{R}$ and codomain $\mathbf{R}$, where

$$f\colon x \mapsto x+1, \quad g\colon \mapsto \sin(x), \quad h\colon x \mapsto x^2.$$

Then $g \circ f(x) = \sin(x+1)$ and $h \circ g(x) = \sin^2(x)$.

Given x, let $y = x+1$, $z = \sin(y)$, $w = z^2$. Then $h \circ g(y) = w$, $g \circ f(x) = z$, and $h(z) = w$. Therefore $(h \circ g)(f(x)) = h(g \circ f(x))$. In other words, $((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$.

It therefore makes sense to write $h \circ g \circ f$, since it evaluates to the same function — that is, of course

$$x \mapsto \sin^2(x+1)$$

— whichever way the parentheses are placed. This distinction is invisible to us, but there are some simple examples of operations which are *not* associative.

Example. Considering addition, subtraction, multiplication, and division over $\mathbf{R}$: which are associative? Which are commutative?

(1.11) Lemma Composition of maps is associative. *If g is compatible with f and h is compatible with g then h is compatible with $g \circ f$ and $h \circ g$ is compatible with f and $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. It is easy to show that range$(g \circ f) \subseteq$ range(g), and we are given range$(g) \subseteq$ domain(h). Therefore range$(g \circ f) \subseteq$ domain(h), so h is compatible with $g \circ f$.

Again, range$(f) \subseteq$ domain$(g) =$ domain$(h \circ g)$, so $h \circ g$ is compatible with f.

For any $x \in$ domain(f), let $y = f(x)$, $z = g(y)$, and $w = h(z)$. See Figure 3.

By definition, $z = g \circ f(x)$, so $w = h(z) = h((g \circ f)(x)) = (h \circ (g \circ f))(x)$.

Again by definition, $w = h \circ g(y) = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x)$.

That is, for every $x \in$ domain(f), $(h \circ (g \circ f))(x) = ((h \circ g) \circ f)(x)$.

Therefore $h \circ (g \circ f) = (h \circ g) \circ f$. Q.E.D.

(1.12) Definition Injective, surjective, and bijective maps. *A map $f\colon A \to B$ is* (i) injective *(or one-to-one) if it maps distinct elements to distinct elementsi. i.e., if $x \neq y$ then $f(x) \neq f(y)$. It is* (ii) surjective *(or onto) if its range equals B, and* (iii) bijective *if it is both injective and surjective.*

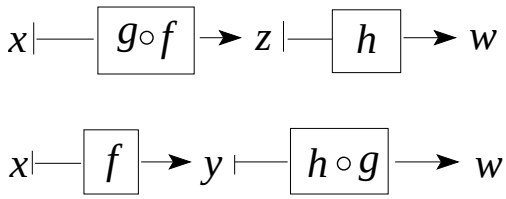


Figure 3: illustrating why $(h \circ (g \circ f))(x) = ((h \circ g) \circ f)(x)$.

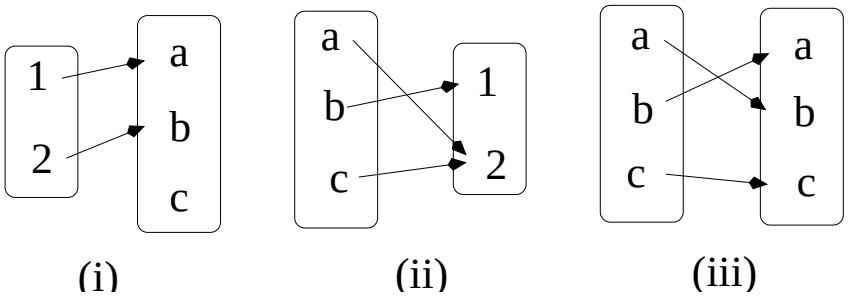


Figure 4: injective, surjective, bijective maps.

(1.13) Lemma (composition of injective maps and surjective maps)..
Given maps f and g, where g is compatible with f,
(i) *if f and g are both injective, so is $g \circ f$.* (ii) *if f and g are both surjective, so is $g \circ f$.* (iii) *If f and g are both bijective, so is $g \circ f$.* (iv) *if $g \circ f$ is injective, so is f* (v) *if $g \circ f$ is surjective, so is g*

Proof. (i) Let x_1 and x_2 be any elements of domain(f). Suppose $x_1 \neq x_2$: R.T.P. $g \circ f(x_1) \neq g \circ f(x_2))$. then $f(x_1) \neq f(x_2)$, so $g(f(x_1)) \neq g(f(x_2))$, as required.
(ii) Let z be any element of the codomain of g. R.T.P. $z = g \circ f(x)$ for some $x \in A$. Choose $y \in B$ such that $g(y) = z$. Then choose $x \in A$ such that $f(x) = y$. Then $g(f(x)) = z$, as required.
(iii) Follows immediately from (i) and (ii).
(iv) Let x_1 and x_2 be any elements of A. Suppose $x_1 \neq x_2$: R.T.P. $f(x_1) \neq f(x_2)$. Since $g(f(x_1)) \neq g(f(x_2))$, $f(x_1) \neq f(x_2)$, as required.
(v) Let z be any element of C. R.T.P. $z = g(y)$ for some $y \in B$. But $z = g(f(x))$ for some $x \in A$, so we can take $y = f(x)$. Q.E.D.

(1.14) Definition (inverse maps). *If $f\colon A \to B$ and $g\colon B \to A$ are two maps such that $g \circ f$ is the identity ι_A on A (1.5), then g is a* left inverse *for f and f a* right inverse *for g. If g is both a left- and right-inverse for f the it is a* two-sided inverse *or simply* inverse *for f.*

Examples. Let $A = [0, \infty)$ (the nonnegative reals. i.e., $A = \{x \in \mathbf{R}\colon x \geq 0\}$) and $B = \mathbf{R}$, $f\colon A \to B$ taking $x \mapsto \sqrt{x}$ and $g\colon B \to A$ taking $x \mapsto x^2$. Then g is a left inverse for f and f a right inverse for g.

(1.15) Lemma *Suppose that $f\colon A \to B$ has a left inverse g and a right inverse h. Then $g = h$ and f has a unique two-sided inverse.*

Proof. Since composition of maps is associative.

$$g \circ (f \circ h) = (g \circ f) \circ h)$$

$$\therefore g \circ \iota_B = \iota_A \circ h$$

$$\therefore g = h$$

If h' is another right inverse, then $g = h'$, so $h = h'$. If g' is another right inverse, then $g' = h$, so $g' = g$. Therefore left and right inverses are unique (and coincide). Since $g = h$, g is a two-sided inverse. Q.E.D.

(1.16) Notation *If $f\colon A \to B$ is a map and Y is any set, then $f^{-1}(Y) = \{x \in A\colon f(x) \in Y\}$. If $y \in Y$, $f^{-1}(y) = f^{-1}(\{y\})$.*

$f^{-1}(Y)$ is called the inverse image of Y under f.

Example. If $f(x) = x^2$, $A = B = \mathbf{R}$, and $Y = \{x \in \mathbf{R}\colon\ -1 \le x \le 4\}$, then $f^{-1}(Y) = \{x \in \mathbf{R}\colon\ -2 \le x \le 2\}$ and $f^{-1}(z) = \{\pm\sqrt{2}\}$.

Remark. $f\colon A \to B$ is surjective if and only if $f^{-1}(y) \neq \emptyset$ for each $y \in B$, and is injective if and only if for each $y \in B$ $f^{-1}(y)$ contains zero or 1 elements.

(1.17) Lemma Existence of inverse maps. *Let $f\colon A \to B$ be a map.*
(i) *If f has a left inverse then it is injective,*
(ii) *If f has a right inverse then it is surjective.*
(iii) *If f is injective and $A \neq \emptyset$ or $b = \emptyset$ then f has a left inverse.*
(iv) *If f is surjective and then f has a right inverse.*
(Technical point: this is equivalent to the so-called Axiom of Choice.)
(v) *f has an inverse iff it is bijective, in which case the inverse is unique.*

Proof. (i) and (ii) follow from 1.13 (iv) and (v) since the identity map ι_A is bijective.

(iii) The case where $B = \emptyset$ is a peculiar special case. If $B = \emptyset$ then $A = \emptyset$ and there exists only one map from A to B, namely, $\iota_\emptyset$. This map is its own inverse.

Otherwise, $A \neq \emptyset$. Choose some element a of A, and define $g\colon B \to A$ as follows.

$$g(y) = \begin{cases} x & \text{if } y = f(x) \text{ for some } x \in A \\ a & \text{if } x \notin \text{range}(f) \end{cases}$$

Since f is injective, this defines $g(y)$ uniquely. Clearly $g(f)x()) = x$ for all $x \in A$, so g is a left inverse for f.

(iv) For each $y \in B$ let $g(y)$ be some element x of $f^{-1}(y)$. This x always exists since f is surjective. Clearly $f \circ g(y) = y$ so g is a right inverse for f.

(v) If f has an inverse then from (i) and (ii) it is bijective. Conversely, if f is bijective, then from (iii) f has a left inverse g (since f is bijective, $A = \emptyset \iff B = \emptyset$), and from (iv), f has a right inverse h. But then $g = h$ is the unique 2-sided inverse for f (Lemma 1.15). Q.E.D.

(1.18) Notation *If $f\colon A \to B$ is bijective, then f^{-1} denotes its unique 2-sided inverse.*

Writing $f^{-1}(y)$ as introduced

(1.19) Permutations of $\{1, \ldots, n\}$. Any bijective map from $\{1, \ldots, n\}$ to itself is called a *permutation of* n *letters*. Small Greek letters like σ are generally used when discussing permutations.

In these lectures we use a simple but non-standard notation: a permutation σ is represented by the arrangement which places $\sigma(i)$ in the i-th position. For now, we say 'permutations act on values, not on position.' Thus, $\sigma = 1342$ maps 1 to itself, 2 to 3, 3 to 4, and 4 to 2. Let $\tau = 2314$. Then $\tau \circ \sigma = 2143$ and $\sigma \circ \tau = 3412$.

It is easy to invert a permutation σ with this notation: just write the identity pernutation underneath, forming an array, sort the columns according to the top values; the inverse is then in the bottom row. For example, $\sigma = 1342$:

$$\begin{array}{cccc} 1 & 3 & 4 & 2 \\ 1 & 2 & 3 & 4 \end{array} \quad \mapsto \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{array}$$

so $\sigma^{-1} = 1423$.

2 Semigroups, monoids, and groups

(2.1) An *associative binary operation* on a set S is a map $f\colon S \times S \to S$ with the property that for all $x, y, z \in S$,

$$f(x, f(y, z)) = f(f(x, y), z)$$

The 'formal' $f(x, y)$ notation for maps is seldom used: rather 'infix form' like $x + y$, $x * y$, $x \cdot y$. The 'binary' refers to the fact that f is a function with two variables.

(2.2) Definition *A* semigroup *consists of a nonempty set S together with an associative binary operation $\cdot$ on S,*

The operation is often called 'multiplication' and somtimes 'addition.'

As in ordinary algebra, the $\cdot$ can be omitted so $x \cdot y$ becomes just xy. (As in ordinary algebra, if the operator symbol is $+$ then it is not omitted.)

(2.3) Examples of semigroups

- $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}, \mathbf{C}$ and $\mathbf{H}$ under addition
- $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}$,**C**and $\mathbf{H}$ under multiplication
- Any nonempty set S with $\cdot$, a strange operation defined as follows: $x \cdot y = y$, for all $x, y \in S$.
- As above, except now $x \cdot y$ is defined as x.
- The set of maps $f\colon X \to X$ where X is any set and the operation is $\circ$, composition of maps.
- The set of 2×2 matrices with coefficients in $\mathbf{N}, \mathbf{Z}, , \mathbf{Q}, \mathbf{R}$, or $\mathbf{C}$ under addition (see below)
- the set of 2×2 matrices ... under multiplication

A 2×2 *matrix* is a square array of 4 numbers enclosed in square brackets, with addition componentwise

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+bg & cf+bh \end{bmatrix}$$

It can be shown easily that matrix addition is associative and commutative. It can be shown more laboriaously that matrix multiplication is associative.

(2.4) Identity and inverses. A *left identity* a in a semigroup $S, \cdot$ is an element such that for all $x \in S$,

$$a \cdot x = x.$$

Similarly, a *right identity* b satisfies $x \cdot b = x$ for all $x \in A$.

A two-sided identity, or *identity* for short, is an element e which is both a left- and a right-identity for S.

(2.5) Lemma *If S possesses a left identity and a right identity they they are equal and S possesses a unique (two-sided) identity. (Easy proof omitted.)*

(2.6) Notational conventions. Often one leaves out the operation sign, so xy might mean $x \cdot y$. However, one usually uses $+$ to represent the operation when the semigroup is commutative, and the addition sign is not omitted. Just like in ordinary algebra.

Generally, a two-sided identity will be represented either as 1 or as e. In a commutative group, where $+$ is the symbol used, the identity is written as 0.

(2.7) Definition *A* monoid *is a semigroup containing a (necessarily unique two-sided) identity.*

(2.8) Definition *Let S be a monoid, e its identity. If a and b are elements of S with $a \cdot b = e$, then a is a* left inverse *for b and b is a* right inverse *for a.*

If b is both a left- and a right-inverse for a the it is called a two-sided inverse *for a or* inverse *for short.*

(2.9) Lemma *Let b be an element of a monoid S. If b has a left inverse a and a right inverse c, then $a = c$ and b has a two-sided inverse which is unique. (Proof same as in Lemma 1.15.)* ∎

(2.10) Definition *A* group *G is a monoid in which every element has an inverse.*

The invrerse of x is usually written x^{-1}, or $-x$ if the group operation is denoted $+$.

3 Groups

Recall a semigroup is a nonempty set together with an associative binary operation on it, a monoid is a semigroup with a 2-sided identity, necessarily unique, and a group is a monoid in which every element has a two-sided inverse, necessarily unique.

In a monoid or group, e or 1 usually denote the identity. In a group, x^{-1} usually denotes the inverse of x.

(3.1) Definition *The* order $|S|$ *of a semigroup, group, or monoid S is the number of elements it contains.*.

There is, essentially, just one semigroup of order 1, and it also happens to be a group. For the only possible operation on $\{a\}$ is $aa = a$.

There are two monoids of order 2, in the sense that given two objects $\{e, a\}$, where e is to be the identity, so $ee = e$ and $ea = ae = e$, we have two choices for aa: either a or e.

	e	a
e	e	a
a	a	a

	e	a
e	e	a
a	a	e

Two see that these tables define associative multiplication, you need to consider all 8 possible values of x, y, z and show in each case that $(xy)z = x(yz)$. In the first table, $x(yz)$ and $(xy)z$ both evaluate to e if $x = y = z = e$, otherwise they both evaluate to a. In the second table, $(xy)z$ and $x(yz)$ both evaluate to a if an odd number of the objects x, y, z is a, otherwise they both evaluate to e.

A quicker way of showing that the operation is associative is to note that the multiplication table for $\{1, 0\}$ has the same form, and that we know to be associative. In the second case, the multiplication table for $\{1, -1\}$ has the same form, and that also is associative. The second one gives a group shince a and e both have inverses. The first is not a group since a does not have an inverse.

(3.2) Lemma *If $G, \cdot$ is a group, then for any $a \in G$, the two maps* (i) $x \mapsto ax$ *and* (ii) $x \mapsto xa$ *are bijective.*

Proof. (i) First, the map $x \mapsto ax$ is injective. If $ax = ay$, then $a^{-1}ax = a^{-1}ay$. We can write these expressions without parentheses because the group operation is associative. Therefore $ex = ey$, so $x = y$, proving that the map is injective. Second, the map is surjective. Given $y \in G$, let $x = a^{-1}y$. Then $ax = aa^{-1}y = y$, proving that the map is surjective. Thus the map is bijective. (ii): almost identical proof. Q.E.D.

Let us count the *essentially different* groups with four elements $\{e, a, b, c\}$. From the above lemma, $ab \neq a$ because $ae = a$, and $ab \neq b$ hecause $eb = b$.

Either $a^2 = e$ or $a^2 \neq e$. In the latter case, we may as well assume that $b = a^2$, otherwise just interchange the roles of b and c.

But if $a^2 = b$, then $ac \neq a, b, c$ so $ac = e$ and $ab = c$. Thus $b = a^2$ and $c = a^3$ and $a^4 = e$. The rest of the multiplacation table is easy to fill in.

Hence we may assume $a^2 = e$. Moreover, if $b^2 \neq e$ or $c^2 \neq e$ then we would be re-inventing the above group with b or c taking the role of a. Hence $a^2 = b^2 = c^2 = e$.

If $a^2 = e$, the $ab \neq a, e, b$ so $ab = c$. Again $ca = b$. This allows the table to be completed. There are only two essentially different groups with four elements (the above happen to be groups, which needs to be checked, only we won't).

	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

These groups happen to be commutativee. In fact the smallest non-commutative group is the group of permutations of 3 letters (its order is $3! = 5$).

4 The symmetric group S_n

(4.1) Lemma *For any set A, the set of bijective maps from A to A is a group under composition.*

Proof. Since composition is associative (REFERENCES) the set is a semigroup. The identity ι_A is a 2-sided identity. Every bijective map $f\colon A \to A$ has a unique 2-sided inverse f^{-1}, so $f \circ f^{-1} = f^{-1}circf = \iota_A$ (REFERENCES) so we indeed have a group. Q.E.D.

(4.2) Definition *Let $n \in \mathbf{N}$. A bijection $f\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a* permutation on n letters. *The group of permutations is called S_n and called the* symmetric group *on n letters.*

(4.3) Notation (nonstandard). Every permutation describes an arrangement, and the permutation $1 \mapsto r, 2 \mapsto s, 3 \mapsto t, \ldots$ is represented by the arrangement $rst \ldots$. In this notation $123 \ldots$ represents the identity permutation.

Example. S_0 contains one element, the empty map. $S_1 = \{1\}$, $S_2 = \{12, 21\}$,
$S_3 = \{123, 132, 213, 231, 312, 321\}$, and
$S_4 = \{1234, 1243, 1324, 1342, 1423, 1432, 2134, 2143, 2314, 2341, 2413,$
$2431, 3124, 3142, 3214, 3241, 3412, 3421, 4123, 4132, 4213, 4231, 4312, 4321\}$.

Position or value? This notation can lead to confusion about composing permutations. For now, we work with values, not positions. Thus, for example,

$$231 \circ 132 = 213$$

This is unnatural when we consider symmetries of the triangle, the 15 puzzle, etcetera, where it is natural for permutations to be defined on position, but for now permutations act on values.

(4.4) Lemma $|S_n| = n!$ *(from school).*

The multiplication table (Cayley table) for S_1 is completely trivial and for S_2 is trivial.

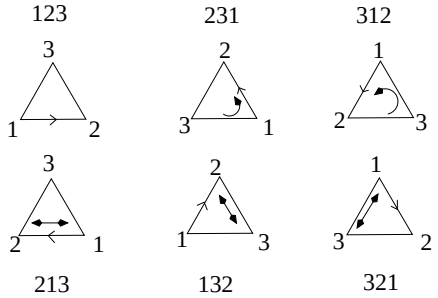
	12	21
12	12	21
21	21	12

For S_3 it is not (this is the smallest noncommutative group).

	123	132	213	231	312	321
123	123	132	213	231	312	321
132	132	123	312	321	213	231
213	213	231	123	132	321	312
231	231	213	321	312	123	132
312	312	321	132	123	231	213
321	321	312	231	213	132	123

(4.5) Symmetries of plane figures. We are interested in a certain group of bijections from the plane to itself, namely, the *rigid linear transformations.* These are maps from the plane to itself which take the origin $(0, 0)$ to itself and preserve distance: rotations and reflections and compositions of rotations and reflections.

Let T be a nonempty set in the plane. The *symmetries of* T are those rigid linear transformations of the plane which map T onto T. The symmetries form a group. We are mostly interested in the case where T is a regular n-sided polygon, such as an equilateral triangle, centred at the origin. The group of symmtries of a regular n-gon ($n \geq 3$) is called the *dihedral group* D_{2n} of order $2n$. Its order is $2n$. To see this, label the corners from 1 to n in anticlockwise order. A symmtry must take the adjacent pair 12 to another adjacent pair $i, i+1$ or $i, i-1$ (where we take $n+1 = 1$ and $1-1 = n$). Different symmetries correspond to different pairs, and each such pair is the image of 12 under a unique symmetry: hence, $2n$ symmetries.



If we label the corners of the n-gon as described, then each symmtery corresponds to a unique permutation of the corners. This way, D_{2n} can be viewed as part of S_n. However, notice that the way symmetries are defined, it makes no sense for the permutations to 'act on values.' They should 'act on positions,' as illustrated. Notice that D_6 corresponds exactly to S_3, and this helps us to understand S_3 better.

5 Generators for S_n

(5.1) Generators of a group. A *set of generators* for a group G is a set $a_1, \ldots, a_p$ of elements of G with the property that every element of G can be expressed as a product of powers (positive or negative) of these elements.

(5.2) Definition *Suppose that $i_1, \ldots, i_k$ is a list of distinct 'letters' in $\{1, \ldots, n\}$. The permutation which takes $i_1 \mapsto i_2$, $i_2 \mapsto i_3$, ..., $i_{k-1} \mapsto i_k$ and $i_k \mapsto i_1$, and leaves all the other 'letters' fixed, is called a* k-cycle.

There is a special notation for this k-cycle: $(i_1 i_2 \ldots i_k)$. It should not be confused with our 'table' notation such as 1342, which is actually a 3-cycle (234).

A 2*-cycle is also called a* transposition.

(5.3) Lemma *Disjoint cycles commute.*

Proof. Let $\sigma = (i_1 i_2 \dots i_k)$ and $\tau = (j_1 j_2 \dots j_\ell)$ be disjoint cycles, that is, the sets i_r and j_s are disjoint.

We shall prove a more general result: if σ and τ together have the property that for all i such that $\sigma(i) \neq i$, $\tau(i) = i$, and for all i such that $\tau(i) \neq i$, $\sigma(i) = i$, then $\sigma\tau = \tau\sigma$..

Note that if $\sigma(i) = j \neq i$ then $\sigma(j) \neq j$, since σ is injective; so $\tau(j) = j$.

R.T.P.: for any letter i, $\sigma\tau(i) = \tau\sigma(i)$.

Either (a) $\sigma(i) = \tau(i) = i$ or (b) $\sigma(i) \neq i$ or (c) $\tau(i) \neq i$.

In case (a), $\tau\sigma(i) = i = \sigma\tau(i)$. In case (b). $\tau(i) = i$ and $\tau(\sigma(i)) = \sigma(i)$, so $\sigma(\tau(i)) = \sigma(i) = \tau(\sigma(i))$. Case (c) is the same as case (b). Q.E.D.

(5.4) Lemma *Every permutation* σ *in* S_n *(except the identity) can be expressed as a product of disjoint cycles.*

Informal Proof. Begin with $i_1 = 1$, let $i_2 = \sigma(i_1)$, $i_3 = \sigma(i_1)$, and continue until an i_k has been found such that $\sigma(i_k) = i_j$ with $j < k$. It is assumed that $\sigma(i_{k-1})$ is not an earlier i_r. Claim $\sigma(i_k) = i_1$.

For if $\sigma(i_k) = i_j$ with $j \geq 2$, then $i_j = \sigma(i_{j-1}$, so $\sigma(i_k) = \sigma(i_{j-1})$. But σ is injective, so $i_k = i_{j-1}$. But $i_k = \sigma(i_{k-1}$, so $\sigma(i_{k-1}) = i_{j-1}$. This contradicts our assumption about k.

It follows that $(i_1 i_2 \dots i_k)$ is a k-cycle σ_1 with the property that $\sigma(i_r) = \sigma_1(i_r)$ for each i_r in the cycle.

If $\sigma_1 = \sigma$ then we are finished. Otherwise choose some j_1 which differs from all the i_r, and construct an ℓ-cycle $\sigma_2 = (j_1 \dots j_\ell)$ where $j_{s+1} = \sigma(j_s)$. It is disjoint from the first cycle σ_1, because if $j_s = i_r$ for some r and s, then $\sigma^{\ell+1-s}(j_s) = \sigma^{\ell+1-s}(i_r)$. But the first is j_1 and the second is another i_t, so $j_1 = i_t$, and j_1 is not left fixed by σ_1.

Continue in this way until σ has been expressed as a product $\sigma_1\sigma_2\cdots$ of disjoint cycles. Q.E.D.

(5.5) Corollary *If* $n \geq 2$ *then every permutation in* S_n *can be expressed as a product of transpositions.*

Proof. In view of Lemma 5.4, It is enough to show that every k-cycle equals a product of transpositions. A 1-cycle (i) (which has no effect) can be written, say, as $(12)(12)$. A 2-cycle is already in the correct form. For $k \geq 3$, a k-cycle can be written as a product of $k-1$ transpositions, as follows.

If $k \geq 3$, consider a k-cycle (a) $(i_1 i_2 \dots i_k)$. Claim that it equals the following product of transpositions

$$\text{(b)} \qquad (i_i i_k)(i_1 i_{k-1}) \dots (i_1 i_3)(i_1 i_2)$$

Given any letter i, if it does not occur in the list $i_1, \dots, i_k$ then heither the k-cycle (a) nor the product (b) affects it. On the other hand, if $j = i_r$ with $1 < r < k$, then the subproduct $(i_1 i_{r+1})(i_1 i_r)$ takes $i_r \mapsto i_1 \mapsto i_{r+1}$, and i_r and i_{r+1} are not mentioned elsewhere in the product (b), so (b) takes i_r to i_{r+1}. If $j = i_k$ then the leftmost transposition $(i_1 i_k)$ in (b) takes i_k to 1. Finally, if $j = i_1$ then the rightmost transposition in (b) takes j to i_2, and on other transposition affects it. Thus (a) and (b) have the same effect. Q.E.D.

In other words: S_n is generated by its transpositions.

(5.6) Definition *In any group G, for any $x \in G$, the map $y \mapsto xyx^{-1}$ is called* conjugation of y by x.

In the Lemma below, we retain the symbol $\circ$ for composition of permutations, because the expression $\sigma(ij)\sigma^{-1}$ suggests that the map σ is applied to ij.

(5.7) Lemma *Given a transposition (ij) and a permutation σ, the result of conjugating (ij) by σ is the transposition $(\sigma(i)\sigma(j))$.*

Proof. If $k \neq \sigma(i), \sigma(j)$, then $(ij) \circ \sigma^{-1}(k) = \sigma^{-1}(k)$, so $\sigma \circ (ij) \circ \sigma^{-1}(k) = k$, On the other hand, if $k = \sigma(i)$, then $(ij) \circ \sigma^{-1}(k) = j$, so $\sigma \circ (ij) \circ \sigma^{-1}(k) = \sigma(j)$. Similarly if $k = \sigma(j)$. Q.E.D.

(5.8) Corollary *Given i, k such that $i + k < n$, conjugating $(i\ i+k)$ by $(i+k\ \ i+k+1)$ yields $(i\ \ i+k+1)$ (trivial).*

(5.9) Corollary *S_n is generated by the two permutations (12) and $(123\ldots n)$.*

Proof. Since the transpositions generate S_n, it is enough to show that every transposition can be generated from (12) and $(12\ldots n)$. Call the n-cycle β.

For $1 \leq i \leq n-1$, $\beta^{i-1}(1) = i+1$ and $\beta^{i-1}(2) = i+2$. It follows that conjugating (12) by β^i yields $(i+1, i+2)$

(The i-th power of β, β^i, has its natural meaning, and the zero-th power is the identity e.) Therefore from (12) and β we can generate the following list of transpositions

$$(12), (23), (34), \ldots (n-1, n)$$

For $1 \leq k \leq n-1$, let T_k be the set of transpositions $\{(i\ \ i+k) \colon 1 \leq i \leq n-k\}$.

The above list of transpositions is T_1. Given a transposition $(i\ \ i+k)$ in T_k, if $i+k < n$, we can conjugate the transpositions by $(i+k\ \ i+k+1)$, which is in T_1, to get $(i\ \ i+k+1)$ (corollary above). In other words, T_{k+1} can be formed from T_k by conjgating by elements of T_1. Now if $(i\ \ i+k)$ and $(i+k\ \ i+k+1)$ are known to be products of powers of (12) and β, then so is the conjugate $(i+k\ \ i+k+1)(i\ \ i+k)(i+k\ \ i+k+1)$.

It follows by induction on k that every transposition in T_k can be generated from (12) an β, for $1 \leq k \leq n-1$. But this includes every transposition (ij) with $1 \leq i < j \leq n$. Since $(ij) = (ji)$, it includes every transposition. Q.E.D.

Example. $(12), (1234)$ generate S_4. Conjugating (12) bey powers of (1234) we get (23) and (34), so we have $(12), (23), (34)$.

Conjugate (12) by (23) to get (13). Conjugate (23) by (34) to get (24). Now we have (13), (24).

Finally conjugate (13) by (34) to get (14). This gives us $(12), (23), (34), (13), (24), (14)$. These are all the 6 transpositions.

6 Parity and the alternating group

We consider the set of all *unordered pairs*

$$\{\{i, j\} \colon 1 \leq i \neq j \leq n\}$$

There are $n(n-1)/2$ such pairs, since $\{i, j\}$ and $\{j, i\}$ are the same. A permutation $\sigma \in S_n$ *inverts* a pair $\{i, j\}$ if ($i < j$ and $\sigma(i) > \sigma(j)$) or ($i > j$ and $\sigma(i) < \sigma(j)$). The *parity* of σ is *even* or *odd* according as σ inverts an even or odd number of these pairs.

(6.1) Lemma *Given a permutation $\sigma \in S_n$, consider the mapping $\{i,j\} \mapsto \{\sigma(i),\sigma(j)\}$. Then this mapping is a bijection of the set of unordered pairs onto itself.*

Proof. Let U denote this set of unordered pairs $\{\{i,j\}\colon 1 \le i,j \le n,\ \ i \ne j\}$.

Given $\{i,j\} \in U$, $\sigma(i) \ne \sigma(j)$, so $\{\sigma(i),\sigma(j)\} \in U$.

To show the map is injective: If $\{\sigma(i),\sigma(j)\} = \{\sigma(k),\sigma(\ell)\}$, then without loss of generality $\sigma(i) = \sigma(k)$ and $\sigma(j) = \sigma(\ell)$. Then $i = k$ and $j = \ell$, so $\{i,j\} = \{k,\ell\}$.

It is surjective: given $\{k,\ell\} \in U$, let $i = \sigma^{-1}(k)$ and $j = \sigma^{-1}(\ell)$. Then $\{k,\ell\} = \{\sigma(i),\sigma(j)\}$. This shows the map is surjective.

Therefore the map is a bijection from U onto itself. Q.E.D.

(6.2) Theorem *Given $\sigma,\tau \in S_n$, if both are even or both or odd then $\tau\sigma$ is even, otherwise $\tau\sigma$ is odd.*

Proof. Let A consist of all pairs $\{i,j\}$ inverted by σ, and let B consist of all pairs $\{i,j\}$ such that τ inverts $\{\sigma(i),\sigma(i)\}$.

The mapping discussed in Lemma 6.1 carries B bijectively onto the set

$$\{\sigma(i),\sigma(j)\}\colon \tau \text{ inverts } \{\sigma(i),\sigma(j)\}\}.$$

Since tbe map is bijective, the latter consists of all pairs inverted by τ.

Thus $|B|$ is the number of pairs inverted by τ.

Consider an ordered pair $\{i,j\}$. Under σ it goes to $\{\sigma(i),\sigma(j)\}$, and under τ that is taken to $\{\tau\sigma(i),\tau\sigma(j)\}$. If in neither A nor B then neither pair is inverted, so $\{i,j\}$ is not inverted by $\tau\sigma$. If it belongs to $A \cap B$ then it is inverted twice, so not invertex by $\tau\sigma$. Otherwise it belongs to $A\backslash B$ or $B\backslash A$ and it is inverted by $\tau\sigma$.

In other words, the total number of inversions of $\tau\sigma$ is $|(A\backslash A \cap B) \cup (B\backslash A \cap B)|$ = $|A| + |B| - 2|A\backslash B|$, which has the same parity as $|A| + |B|$. But $|A|$ and $|B|$ are the number of pairs inverted by σ and τ respectively. Q.E.D.

(6.3) Corollary *The identity permutation is even, and σ and σ^{-1} always have the same parity.*

Proof. The identity permutation e inverts no pairs, so it is even. Since $\sigma\sigma^{-1} = e$, the sum of parities of σ and σ^{-1} is even, so both have the same parity. Q.E.D.

(6.4) Definition *A* subgroup *of a group G is a nonempty sybset H which is closed under multiplication and inversion. It is a group in its own right.*

(6.5) Corollary *The set of even permutations in S_n is a subgroup.*

Proof. It contains the identity e; the product of two even permutations is even, so it is closed under multiplication; and the inverse of an even permutation is even, so it is closed under inversion. Q.E.D.

(6.6) Definition *The subgroup of even permutations in S_n is called the* alternating group *and written A_n.*

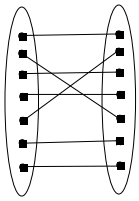


Figure 5: where lines cross, a pair is inverted. Hence transpositions are odd.

(6.7) Lemma *Transpositions are odd.*

Proof. If we represent a permutation by a diagram, then where lines cross we have an inverted pair: see Figure 5. Let (ij) be a transposition, where w.l.o.g. (without loss of generality), $i < j$.

Let $\alpha = (ij)$. Consider a pair $\{k, \ell\}$ inverted by α. Certainly one of k and ℓ must be i or j.

Suppose $k \neq i, j$. Since $k < \ell$, $k > \alpha(\ell)$. Hence $\ell = j$ and $k > i$. Similarly if $\ell \neq i, j$, then $k = i$ and $\ell < j$. Therefore each k, $i < k < j$, is involved in two inverted pairs, $\{i, k\}$ and $\{k, j\}$. The only remaining possibility is $\{k, \ell\} = \{i, j\}$, which is inverted. There are a total of $2(j - 1 - 1) + 1$ pairs inverted by (ij), which is therefore odd. Q.E.D.

(6.8) Lemma *A k-cycle is even if k is odd and vice-versa.*

Proof. 1-cycles represent the identity permutation which is even. for $k \geq 2$, any k-cycle can be expressed as a product of $k - 1$ transpositions, each of which is odd. Hence if k is odd the cycle is even and vice-versa. Q.E.D.

(6.9) Corollary *If σ is expressed as a product of (disjoint) cycles, its parity is the parity of the number of even-length cycles in this product. (Proof immediate.)*

(6.10) Lemma (i) *Every cycle of odd length $k \geq 3$ is expressible as a product of 3-cycles.* (ii) *Every cycle of even length ≥ 4 can be written as a product of 3-cycles multiplied either on the left r on the right by a transposition.* (iii) *Every product of two even-length cycles can be written as a product of 3-cycles, only affecting those letters involved in the cycles.*

Proof. (i) $(i_1 i_2 \dots i_k) = (i_1 i_k)(i_1 i_{k-1}) \cdots (i_1 i_3)(i_1 i_2)$. If k is odd, The transpositions can be grouped in pairs, giving $(i_1 i_{k-1} i_k)(i_1 i_{k-3} i_{k-2}) \cdots (i_1 i_2 i_3)$.

(ii) If k is even, all but one transposition can be grouped in pairs, leaving one on the left or one on the right. We get either $(i_1 i_k)(i_1 i_{k-2} i_{k-1}) \cdots (i_1 i_2 i_3)$ or $(i_1 i_{k-1} i_k) \cdots (i_1 i_3 i_4)(k_1 i_2)$.

(iii) Express the first as $P(ij)$ and the second as $(k\ell)Q$, where P and Q are either e or a product of 3-cycles. Now $(ij)(k\ell) = (ij)(ik)(ik)(k\ell) = (ikj)(k\ell i)$, so the product can be written as $P(ikj)(i\ell k)Q$. Q.E.D.

(6.11) Corollary *If $n \geq 3$ then the 3-cycles in S_n generate A_n.*

Proof. Let σ be an even permutation in S_n, $\sigma \neq e$. It can be expressed as a product of disjoint cycles of length ≥ 2, of which an even number are of even length. The odd-length cycles can be writted as products of 3-cycles. The even-length cycles can be taken in pairs, in any order, and each pair can be expressed as a product of 3-cycles from part (iii) of the above Lemma. Q.E.D.

7 Binary relations

The aim of this section is to exhibit the correspondence between equivalence relations on a set X and partitions of X (Theorem 7.13).

A binary relation R on a set X is a property of pairs of elements of X. If R is a binary relation on X and $x, y \in X$ we usually write xRy to mean that x and y are in the relation R.

This notation is familiar, e.g., $x \neq y$.

Technically, R is specified completely by the set $\{(x, y) \in X \times X \colon xRy\}$. Therefore a binary relation on X is defined as any subset of $X \times X$.

(7.1) Examples of binary relations

- The equality relation $=$ on any set.
- The inequality relation $\neq$ on any set.
- The relation $<$ on $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$,
- The relation $\leq$ on $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$,
- The relation $\subseteq$ on any set of sets.
- The relation 'i and j have the same parity' on $\mathbf{N}$ or on $\mathbf{Z}$.
- Given any integer n, the relation 'n divides $x - y$', also written '$x \equiv y \pmod{n}$' on $\mathbf{Z}$.
- Given any group G and a subgroup H of G, the relation '$x^{-1}y \in H$.'
- Given G and H, the relation '$xy^{-1} \in H$.' This is not necessarily the same relation as the preceding.

(7.2) Definition *A binary relation R on X is*

- **reflexive** *if for all $x \in X$, xRx*
- **symmetric** *if for all $x, y \in X$, $xRy \Rightarrow yRx$.*
- **transitive** *if for all x, y, z in X, if xRy and yRz then xRz.*

An equivalence relation *is one which is reflexive, symmetric, and transitive.*

(7.3) The equality relation $=$ on any set is an equivalence relation. Broadly speaking, an equivalence relation has the important properties of an equality relation, and can be treated as equality. For example,

(7.4) Lemma *The rational number system $\mathbf{Q}$ is $\{x/y \colon x, y \in \mathbf{Z} \ \& \ y \neq 0\}$. The relation $=$ on $\mathbf{Q}$, where $a/b = c/d$ if $ad = bc$, is an equivalence relation.*

Proof. Reflexivity: i.e., $a/b = a/b$ since $ab = ba$.
Symmetry: if $a/b = c/d$, then $ad = bc$. RTP: $cb = da$. But $da = ad$ (integer multiplication is commutative) and $ad = bc$ so $da = bc$ (= is transitive on **Z**). Also $bc = cb$, so $da = cb$. Therefore $cb = da$ (= is symmetric on **Z**).
Transitivity: suppose $a/b = c/d$ and $c/d = e/f$. Then $ad = bc$ and $cf = de$. RTP $af = be$. Since $ad = bc$, $adf = bcf$. But $cf = de$, so $bcf = bde$. Therefore $adf = bde$. But $d \neq 0$, so it can be cancelled out (property of integer multiplication), and $af = be$. Q.E.D.

This 'equality' relation on fractions is looser than complete identity: $1/2 = 6/12 = (-2)/(-4)$.

(7.5) Definition *Let R be an equivalence relation on a set X. For any $x \in X$ the* equivalence class of x modulo R, *or $[x]_R$ for short, is the set*

$$\{y \in X : yRx\}$$

Sometimes we write just $[x]$ if it is obvious what R is.

(7.6) Let R be the relation on **Z**: xRy iff $x - y$ is even. This is an equivalence relation. There are two equivalence classes: the even integers, and the odd integers.

(7.7) Definition *A* partition P *of a set X is a family of pairwise disjoint, nonempty, subsets of X whose union is X.*

(7.8) Lemma *Let R be an equivalence relation on X. Given $x, y \in X$, xRy iff $[x]_R = [y]_R$.*

Proof: If: suppose $[x]_R = [y]_R$. Then $x \in \{z : zRy\}$, so xRy.
Only if: suppose xRy. If $z \in [x]_R$, zRx, so zRy by transitivity, so $z \in [y]_R$. Therefore $[x]_R \subseteq [y]_R$. Again yRx so by the same reasoning $[y]_R \subseteq [x]_R$. Therefore $[x]_R = [y]_R$. Q.E.D.

(7.9) Corollary *If R is an equivalence relation on X then its equivalence classes form a partition of X.*

Proof. For each x in X, $x \in [x]_R$ because xRx, so the equivalence classes are nonempty, and every x belongs to at least one equivalence class, so the union is X. It remains to prove that the classes are pairwise disjoint, i.e., if $[x]_R \cap [y]_R \neq \emptyset$ then $[x]_R = [y]_R$.
Suppose $[x]_R \cap [y]_R \neq \emptyset$. Choose $w \in [x]_R \cap [y]_R$. By definition wRx, so xRw, and wRy, so xRy by transitivity. Therefore $[x]_R = [y]_R$ (Lemma 7.8), as required. Q.E.D.
Equivalence relation defined from a partition. Let P be a partition of X. For any $x \in X$, let P^x be the unique set $C \in P$ such that $x \in C$. (The notation P^x is only used in this section.)

(7.10) Lemma *Let P be a partition of X, and $x, y \in X$. The following are equivalent.*
(i) $P^x = P^y$, (ii) $x \in P^y$, *and* (iii) *for some $C \in P$, x and y both belong to C.*

Proof. (i) implies (ii), since $x \in P^x$. Assuming (ii), it follows that x and y both belong ot P^y, so (iii) holds. Therefore (ii) implies (iii). Assuming (iii), $C = P^x = P^y$, so (i) holds. Therefore (iii) implies (i). Q.E.D.

(7.11) Lemma *With X and P as above, let xRy mean that x and y satisfy* (i), (ii), *or* (iii) *in the above lemma. Then* (i) *R is an equivalence relation, and* (ii) *P is the set of its its equivalence classes.*

Proof. (i) Reflexivity: $P^x = P^x$. Symmetry: $P^x = P^y$ implies $P^y = P^x$. Transitivity: Suppose $P^x = P^y$ and $P^y = P^z$. Then $P^x = P^z$. Therefore R is an equivalence relation.

(ii) For any $y \in X$, $[y]_R = \{x \in X : x \in P^y\}$, using formulation (ii) of Lemma 7.10 $= P^y$, so $[y]_R \in P$. For any $C \in P$, $C \neq \emptyset$: choose $y \in C$. Then $C = P^y$ by definition, so $C = [y]_R$ is an equivalence class of R. Q.E.D.

(7.12) Lemma *Let $\sim$ be an equivalence relation on X, let P be the set of its equivalence classes, and let R be defined from P as in Lemma 7.11. Then the relations R and $\sim$ are identical.*

Proof. For any $y \in X$, $P^y = [y]_\sim$. Therefore

$$xRy \Leftrightarrow x \in P^y \Leftrightarrow x \in [y]_\sim \Leftrightarrow x \sim y,$$

Q.E.D.

Summarising these results.

(7.13) Theorem *Let X be a set. For any equivalence relation $\sim$ on X, let $C(\sim)$ be the set of its equivalence classes. For any partition P of X let $R(P)$ be the relation $P^x = P^y$, $x \in P^y$, or x and y both helong to the same set in P. All three versions are equivalent (Lemma 7.10).*

Then $C(\sim)$ is a partition of X and $R(P)$ is an equivalence relation on X. Also, $C(R(P)) = P$ and $R(C(\sim)) = \sim$.

Thus C and R are mutually inverse one-to-one correspondences between the equivalence relations on X and partitions of X.

In short: there is a one-to-one correspondence between equivalence relations on X and partitions of X.

Proof. By lemma 7.9 $C(\sim)$ is a partition of X. By Lemma 7.11, $R(P)$ is an equivalence relation on X. By part (ii) of the same lemma, $C(R(P)) = P$. Finally by Lemma 7.12, $R(C(\sim)) = \sim$. Q.E.D.

8 Remainder modulo n and integer division

(8.1) Definition *Given integers m and n, n* divides *m, written $n|m$, if there exists an integer q such that $m = qn$.*

(8.2) Definition *Let n be a positive integer. The relation $x \equiv y$ (mod n) means $n|(x - y)$.*

(8.3) Lemma *This is an equivalence relation.*

Proof. Reflexivity: $n|(x - x)$. Symmetry: if $n|(x - y)$, so $x - y = qn$ for some q, then $y - x = (-q)n$ so $n|(y - x)$.

Transivity: if $n|(x - y)$ and $n|(y - z)$, then $x - y = q_1 n$ and $y - z = q_2 n$ for some $q_1, q_2 \in \mathbf{Z}$. Then $x - z = (q_1 + q_2)n$ so $n|(x - z)$. Q.E.D.

In the definition below, we invoke a very useful property of the natural number system **N**, called the *principle of well-orering*, or PWO for short:

For every nonempty subset S of $\mathbf{N}$, S contains a smallest element s. In other words, there exists a nonnegative integer s such that $s \in S$, and for any other $r \in \mathbf{N}$, if $r < s$ then $r \notin S$.

This principle is very close to the principle of Mathematical Induction. It is useful for defining certain natural numbers or certain mappings into $\mathbf{N}$.

(8.4) Definition *Let n be a positive integer, x any integer. The* remainder *(or residue)* of x modulo n *or $x \mod n$ is the smallest integer in the equivalence class of x (mod n), i.e., the smallest element of $[x]_{(\equiv \mod n)} \cap \mathbf{N}$.*

(8.5) Lemma *For all $x \in \mathbf{Z}$,* (i) $x \mod n$ *is a well-defined nonnegative integer, and* (ii) $0 \leq x \mod n \leq n-1$.

Proof. The definition invvkes the well-ordering principle (36.1). We need show that there exists a nonnegative integer s such that $n|x-s$. Then the PWO is being used correctly, and (i) follows.

In other words, we need to exhibit an s such that $x - s = qn$ for some $q \in \mathbf{Z}$ Then $s = x - qn$. If $x \geq 0$ we can take $q = 0$ so $s = x$. If $x < 0$ take $q = x$. Then $x - qn = x - xn = x(1-n)$. Since $x < 0$ and $1 - n \leq 0$, $x(1-n) \geq 0$ as required for (i): $x \mod n$ is well defined.

(ii) Let $r = (x \mod n)$, the smallest $r \geq 0$ such that $n|(x-r)$. Since $r \geq 0$, it remains to show that $r < n$, or $r - n < 0$.

Note that $n|(x - r + n)$ also, i.e., $n|(x-(r-n))$. Since $r - n < r$, it must be that $r - n < 0$, so $r < n$. i.e., $r \leq n - 1$. Q.E.D.

(8.6) Lemma $x \mod n$ *is the unique $r \in \{0, \ldots, n-1\}$ such that $n|(x-r)$.*

Proof. Let $r = x \mod n$, so $n|(x-r)$ and $0 \leq r \leq n-1$. Suppose that $n|(x-r')$ where $r' \neq r$ and $r' \geq 0$. RTP: $r' \geq n$. By minimality of r, $r' > r$. Since $n|(x-r)$ and $n|(x-r')$, $n|(r'-r)$, so $r' - r = qn$ for some $q \in \mathbf{Z}$. Since $r' - r > 0$ and $n > 0$, $q > 0$ and $qn \geq n$, so $r' \geq r + n \geq n$. Q.E.D.

We shall use the result below, without discussing it

(8.7) Lemma (cancellativity of Z). *Given $q, q', n \in \mathbf{Z}$ where $n \neq 0$, if $qn = q'n$ then $q = q'$.* ∎

(8.8) Lemma *Given $x \in \mathbf{Z}$ and a positive integer n, there exist unique q and r such that $x = qn + r$ with $0 \leq r \leq n-1$.*

Proof. Existence: let $r = (x \mod n)$ so $n|x-r$, i.e., $x - r = qn$ for some q, so $x = qn + r$ and $0 \leq r \leq n-1$.

Uniqueness: Suppose that $x = qn + r = q'n + r'$ where $0 \leq r, r' \leq n-1$. We already know $r = r' = (x \mod n)$, so $qn = q'n$, and $n \neq 0$, so $q = q'$ by the cancellativity property. Q.E.D.

(8.9) Definition *Given a positive integer n and an integer x, the unique q and r such that $x = qn + r$ where q is an integer and r an integer between 0 and $n-1$, are called the* quotient *and* remainder, *respectively, on dividing x by n. The remainder is, of course, $x \mod n$.*

We write $x \div n$ for the quotient.

(8.10) Lemma *If* $x_1 \equiv y_1 \mod n$ *and* $x_2 \equiv y_2 \mod n$*, then* $x_1 + x_2 \equiv y_1 + y_2 \mod n$*. (Proof omitted.)* **Note: this may have been stated incorrectly in class.**

(8.11) Lemma *Given* $x, y, z \in \mathbf{Z}$ *and* $n > 0 \in \mathbf{Z}$*,*

$$((((x + y) \mod n) + z) \mod n) = ((x + ((y + z) \mod n)) \mod n)$$

(Proof omitted. It would use Lemma 8.10.)

(8.12) $\mathbf{Z}_n$, the additive group of integers modulo n. Given a positive integer n, $\mathbf{Z}_n$ is defined as follows: its elements are $\{0, \ldots, n-1\}$. and its operation is addition (modulo n).

(8.13) Lemma *Z_n is an additive group of order n.*

Proof. The operation is associative from the Lemma 8.11. 0 is the additive identity, $n - x$ is the inverse of x, and the group is commutative because $x + y = y + x$, so $((x + y) \mod n) = ((y + x) \mod n)$. Q.E.D.

For example, $\mathbf{Z}_4$ is

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

9 Additive subgroups of Z

(9.1) Definition *A commutative group is called* abelian. *Usually one uses '$+$' to denote the group operation, its identity is called* 0*, and the inverse of* x *is denoted* $-x$*.*

(9.2) Definition *In a group G, the subgroup generated by elements $a, b, c \ldots$ is denoted $\langle a, b, c, \ldots \rangle$.*
It consists of all elements which can be formed as products of powers of $a, b, c, \ldots$.
If the group is abelian then

$$\langle a, b, c, \ldots \rangle = \{ra + sb + tc + \ldots : r, s, t, \ldots \in \mathbf{Z}\}$$

Remark. 'Products of powers' includes negative powers, where a^{-k} is defined as $(a^{-1})^k$. If the group is not abelian, we cannot group the powers of a together and the powers of b togeter and so on. For examplea $S_3 = \langle (12), (13) \rangle$ because $(23) = (12)(13)(12)$: the transposition (12) occurs twice in the expression.

(9.3) Lemma *Suppose that a, b are elements of a subgroup H of G. Then $\langle a, b \rangle$ is a subgroup of H.*

Proof. Elements of $\langle a, b \rangle$ are products of powers of a and b. Since H is a subgroup it is closed under formation of powers and of products, so every element of $\langle a, b \rangle$ also belongs to H. Q.E.D.

(9.4) Lemma *Given $a, b \in \mathbf{Z}$ where $b > 0$,* $(a \mod b) \in \langle a, b \rangle$.

Proof. Write $a = qb + r$ where $r = a \mod b$. Therefore $r = a - qb$ which is in $\langle a, b\rangle$. Q.E.D.

(9.5) Theorem *Every subgroup of* **Z** *is generated by a unique nonnegative integer* a*; thus the only subgroups of* **Z** *are of the form* $\langle a\rangle$ *for some unique nonnegative integer* a*.*

Proof. The 'trivial' subgroup $\{0\}$ contains only 0.

Let H be any nontrivial subgroup. H contains some nonzero element x; if $x < 0$ then $-x \in H$ also; so H contains a positive integer. By The PWO (principle of well-ordering) it contains a least positive integer n.

Let x be any element of H, so $\langle (x \mod n)n\rangle = \langle x, n\rangle$ is a subgroup of H (Lemma 9.3). so $(x \mod n) \in H$. It is nonnegative and less than n, so it cannot be positive by definition of n, so it is zero.

Since $x = qn + (x \mod n)$ for some q, $x = qn$ for some q. Therefore every element of H is a multiple of n, so $H \subseteq \langle n\rangle$. Since $n \in H$, $\langle n\rangle \subseteq H$, so $H = \langle n\rangle$.

Finally, n is unique. Certainly, $\{0\}$ is generated by 0 and no other integer. If $H = \langle n_1\rangle = \langle n_2\rangle$ where n_1 and n_2 are nonntegative, if H is nontrivial, then both n_1 and n_2 are positive, and they are multiples of each other: $n_1 = q_1 n_2$ and $n_2 = q_2 n_1$, where $q_1, q_2 > 0$.. Then $n_1 = q_1 q_2 n_1$, so $q_1 q_2 = 1$ by the cancellativity, which is only possible if $q_1 = q_2 = 1$. Q.E.D.

Example. A little experimentation will show that $\langle 12, 8\rangle = \langle 4\rangle$.

10 Greatest common divisor

(10.1) Definition *Given* $a, b \in$ **Z***, not both zero,* $\gcd(a, b)$ *is the unique positive generator of the subgroup* $\langle a, b\rangle$ *of* **Z***.*

'Gcd,' or greatest common divisor, means the same as 'hcf,' highest common factor.

(10.2) Corollary $\gcd(a, b)$ *is the largest positive integer which divides both* a *and* b*.*

Proof. Write g for $\gcd(a, b)$, so $\langle a, b\rangle = \langle g\rangle$. Since $a, b \in \langle g\rangle$, $g|a$ and $g|b$.

Since $g \in \langle a, b\rangle$, $g = ra + sb$ for some $r, s \in$ **Z**, if d is a positive integer and $d|a$ and $d|b$ then $d|ra + sb$, i.e., $d|g$, so $d \le g$. Therefore g is the largest integer which divides both a and b. Q.E.D.

The definition of $\gcd$ gives no way of computing it. There is an efficient algorithm, due to Euclid, based on the following lemma.

(10.3) Lemma *Given* $a, b \in$ **Z** *where* $b > 0$*,* $\langle a, b\rangle = \langle (a \mod b), b\rangle$*.*

Proof. Write $a = qb + r$ where $r = a \mod b$. An element of $\langle a, b\rangle$ has the form $sa + tb$. Substitute $qb + r$ for a: $s(qb + r) + tb = sr + (qs + t)b$, which shows it belongs to $\langle r, b\rangle$.

An element of $\langle r, b\rangle$ has the form $sr + tb$. Substitute $(-qb) + a$ for r: $s(-qb + a) + tb = sa + (t - qs)b$. This shows it belongs to $\langle a, b\rangle$.

Therefore the two subgroups are the same. Q.E.D.

(10.4) Corollary *If* $y > 0 \in$ **Z***, then* $\gcd(x, y) = gcd(y, x \mod y)$*.*

Proof. We know that $\langle x, y\rangle = \langle y, (x \mod y)\rangle$. Since the positive generator of $\langle x, y\rangle = \langle y, (x \mod y)\rangle$ is unique (Theorem 9.5), $\gcd(x, y) = \gcd(y, x \mod y)$. Q.E.D.

Euclid's gcd algorithm is based on the above corollary.

```
To calculate gcd(a,b), assuming
a is positive, b nonnegative, and
a >= b.

x := a
y := b

Do the following while y > 0:
    z := x mod y
    x := y
    y := z
----------------------------

When y = 0, x will equal gcd (a,b)
```

For example, to compute $\gcd(1625, 299)$.

x	y	z
1625	299	130
299	130	39
130	39	13
39	13	0

This algorithm can be enhanced to produce integers r and s such that $ra = sb = \gcd(a, b)$. Let us do the calculations as follows:

$x = 1625, y = 299, q = 1625 \div 299 = 5, z = x - qy = a - 5b$.
$x = 299, y = 130, q = 299 \div 130 = 2, z = x - 2y = b - 2(a - 5b) = -2a + 11b$
$x = 130, y = 39, q = 130 \div 39 = 3, z = x - 3y = (a - 5b) - 3(-2a + 11b) = 7a - 38b$
$x = 39, y = 13, z = 0$

Thus $13 = 7(1625) - 38(299)$.

(10.5) Euclid's enhanced gcd algorithm, in tabular form. The calculations can be included in a table as follows. The table includes extra columns for q, the quotient, $r_x, s_x, r_y, s_y, r_z, s_z$, where $x = r_x a + s_x b$, etcetera. Initially, $x = a$, so $r_x = 1, s_x = 0$, and $y = b$, so $r_y = 0, s_y = 1$. Generally, when $q = x \div y$, $z = x \mod y$, so $z = x - qy$, $r_z = r_x - qr_y$ and $s_z = s_x - qs_y$.

x	y	z	q	r_x	s_x	r_y	s_y	r_z	s_z
1625	299	130	5	1	0	0	1	1	-5
299	130	39	2	0	1	1	-5	-2	11
130	39	13	3	1	-5	-2	11	7	-38
39	13	0	3	-2	11	7	-38	$--$	$--$
13	0	$--$	$--$	7	-38	$--$	$--$	$--$	$--$

11 Multiplicative group $\mathbf{Z}_n^*$

Multiplicative properties of congruence (mod n).

(11.1) Lemma *If $x_1 \equiv y_1 \mod n$ and $x_2 \equiv y_2 \mod n$ then $x_1x_2 \equiv y_1y_2 \mod n$. (Proof omitted.)* ∎

(11.2) Corollary $((xy \mod n)z \mod n) = (x(yz \mod n) \mod n)$ *(Proof omitted.)*

(11.3) Corollary *If $n \geq 2 \in \mathbf{Z}$ then the set $\{0, \ldots, n\}$ is a commutative monoid under multiplication* $\mod n$.

Proof. The set is closed under multiplication mod n, which is associative by the above lemma. Clearly $xy \mod n = yx \mod n$, so it is commutative. The multiplicative identity is 1. Q.E.D.

(11.4) Definition *Two integers a, b are* relatively prime *if* $\gcd(a, b) = 1$.

Note: $\gcd(a, b) = \gcd(b, a)$ obviously.

(11.5) Lemma *If a is relatively prime both to b and to c, then a is relatively prime to the product bc.*

Proof. $\gcd(a, b) = ra + sb$, say, and $\gcd(a, c) = ta + uc$, say. Thus $ra + sb = ta + uc = 1$. R.T.P. $\gcd(a, bc) = 1$. Multiply:

$$rta^2 + stab + ruac + subc = 1; \quad r'a + s'bc = (rta + stb + ruc)a + (su)bc = 1$$

So we have integers r', s' such that $r'a + s'bc = 1$. $\gcd(a, bc)$ divides $r'a + s'bc$, so $\gcd(a, bc)$ = 1. Q.E.D.

(11.6) Corollary *Let $\mathbf{Z}_n^*$ be the set of integers in $\{0, \ldots, n\}$ which are relatively prime to n. Then $\mathbf{Z}_n^*$ is an abelian group under multiplication ($\mod n$).*

Proof. It is commutative since it is part of a commutative monoid.

Suppose $0 \leq a, b \leq n - 1$ and $\gcd(n, a) = \gcd(n, b) = 1$. (So $a, b \neq 0$). According to the above lemma, $gcd(n, ab) = 1$. But $\gcd(n, ab \mod n) = \gcd(n, ab)$ (Corollary 10.4), so $gcd(n, ab \mod n) = 1$, so $\mathbf{Z}_n^*$ is closed under multiplication mod n.

Since $\gcd(n, 1) = 1$ (obviously), $1 \in \mathbf{Z}_n^*$.

Given $a \in \mathbf{Z}_n^*$, choose $r, s \in \mathbf{Z}$ so that $ra + sn = 1$. Let $b = r \mod n$. Suppose that $r = qn + b$. Then $ba \mod n = (r - qn)a \mod n = ra \mod n$, because ra and $(r - qn)a$ are congruent mod

n so they have the same remainder mod n. But $ra \mod n = ra + sn \mod n = 1$. Therefore b is the multiplicative inverse of a, so $\mathbf{Z}_n^*$ is a group. Q.E.D.

For example, $\mathbf{Z}_9^* = \{1, 2, 4, 5, 7, 8\}$. Its cayley table is

	1	2	4	5	7	8
1	1	2	4	5	7	8
2	2	4	8	1	5	7
4	4	8	7	2	1	5
5	5	1	2	7	8	4
7	7	5	1	8	4	2
8	8	7	5	4	2	1

12 Cosets, Lagrange's Theorem, and Fermat's Theorem

Given a subgroup H of a group G, it has already been mentioned that the two relations

$$x^{-1}y \in H, \qquad xy^{-1} \in H$$

are both equivalence relations, not necessarily the same relation. The equivalence classes of the first relation are of the form

$$\begin{gathered}[x] = \{y\colon x^{-1}y \in H\} = \\ \{y\colon x^{-1}y = h\} \text{ for some } h \in H = \\ \{y\colon y = xh\} \text{ for some } h \in H = \\ \{xh\colon h \in H\}\end{gathered}$$

The set $\{xh\colon h \in H\}$ is also written as

$$xH$$

and called a *left coset* of H.

For a simple example, let $G = \mathbf{Z}$ and let $H = \langle 2\rangle$, the even integers. The relation is just $y - x \in \langle 2\rangle$, which is just the relation $x \equiv y \mod 2$. There are two distinct (left) cosets, $[0] = \langle 2\rangle$, the even integers, and $[1] = 1 + \langle 2\rangle$, the odd integers.

Considering the second relation, $xy^{-1} \in H$, By the same kind of calculation, the equivalence classes will be seen to be

$$[x] = \{hx\colon h \in H\} = Hx.$$

Hx is called a *right coset* of H. The left- and right cosets of H are sometimes the same, sometimes not. They are the same in an abelian group.

(12.1) Lemma *All left- (and right-) cosets of H in G have the same cardinality, namely, $|H|$.*

Proof. We shall only consider left-cosets. The proof for right-cosets is much the same. Given $g \in G$, R.T.P. $|gH| = |H|$. Consider the following map

$$G \to G\colon \quad x \mapsto gx.$$

THe map $x \mapsto g^{-1}x$ is a two-sided inverse:

$$x \mapsto gx \mapsto g^{-1}(gx) = x, \qquad x \mapsto g^{-1}x \mapsto g(g^{-1}x) = x.$$

Therefore these maps are bijections. The image of H under the map $x \mapsto gx$ is $\{gh\colon h \in H\} = gH$, a left coset. Since bijective maps preserve cardinality (this is gone into in the appendix, Section 37), $|H| = |gH|$. Q.E.D.

(12.2) Corollary (Lagrange's Theorem). *Suppose H is a subgroup of a finite group G, and suppose it has k left cosets. Then $|G| = k|H|$, so $|H|$ divides $|G|$.*

Proof. Choose $g_1, g_2, \ldots g_k$, elements of the k distinct left cosets, so

$$G = g_1H \cup g_2H \cup \ldots g_kH.$$

These cosets form a partition of G, so

$$|G| = |g_1H| + |g_2H| + \ldots |g_kH|.$$

that is, from the above Lemma, $|G| = k|H|$. Q.E.D.

(12.3) Definition *Given an element x of a group G, the* order of x, *or $|x|$, is the smallest positive integer m such that $x^m = e$, if it exists. If $x^m \neq e$ for all positive integers m, then $|x| = \infty$.*

(12.4) Lemma $|x| = |\langle x\rangle|$.

Proof. We consider two cases.

Case (a): suppose that all the nonnegative powers of x, x^r, $(r \in \mathbf{N})$, are distinct. Then $\langle x\rangle$ contains infinitely many distinct elements, so $|\langle x\rangle| = \infty$. Also, $|x| = \infty$, since $x^r \neq x^0$ for all positive integers r. Hence $|x| = |\langle x\rangle| = \infty$ in this case.

Case (b): suppose that there exist $0 \leq r < r + m \in \mathbf{N}$ such that $x^r = x^{r+m}$. Then multiplying by x^{-r}, we get $x^0 = x^m$, i.e., $x^m = e$. Choosing m as small as possible, $m = |x| < \infty$ in this case.

What the remaining argument shows is the subgroup $\langle x\rangle$ is essentially the additive group $\mathbf{Z}_m$.

The powers $x^0, x^1, \ldots, x^{m-1}$ are distinct, because if $x^i = x^j$, $0 \leq i < j \leq m-1$, then $x^{j-i} = e$ and $j - i < m$, a contradiction. Given any $k \in \mathbf{Z}$, $k = qm + r$, where $0 \leq r < m$, and $x^k = x^{qm+r} = (x^m)^q x^r = x^r$. Therefore, $\langle x\rangle = \{e, x, x^2, \ldots x^{m-1}$, and $m = |x| = |\langle x\rangle|$. Q.E.D.

(12.5) Corollary *If $|G| = n < \infty$ and $x \in G$ then $x^n = e$.*

Proof. $|x| = |\langle x\rangle|$ and the latter is finite because G is finite. Write $|x| = m$. By Lagrange's Theorem $|\langle x\rangle|$ divides $|G|$. Therefore $n = qm$ for some integer q, and $x^n = (x^m)^q = e$. Q.E.D.

We can deduce Fermat's Little Theorem from this.

(12.6) Definition *If $x|y \in \mathbf{Z}$ then we say that x is a* factor *of y. A* prime number *is an integer $p \geq 2$ whose only factors are 1 and p.*

(12.7) Lemma *If p is prime and $q \in \mathbf{Z}$ and $\gcd(p,a) \neq 1$, then $\gcd(p,a) = a$ and $p|a$.*

Proof. $\gcd(p,a)$ is a factor of p, so if it is greater than 1 then $\gcd(p,a) = p$. Since $\gcd(p,a)|a$, $p|a$. Q.E.D.

(12.8) Corollary *If p is prime then $|\mathbf{Z}_p^*| = p-1$.*

Proof. Given $1 \leq a \leq p-1$, then $a \mod p \neq 0$, so $p \not| a$, so $\gcd(p,a) = 1$. This means that the group Z_p^* (Corollary 11.6) contains the elements $\{1, \ldots, p-1\}$, a set of cardinality $p-1$. Q.E.D.

There is a technical difficulty in deducing Fermat's Little Theorem from the above corollary. It is 'swept under the carpet' in the following lemma.[1]

(12.9) Lemma (awkward). *Suppose $y \in \mathbf{Z}_n^*$ and $x \in \mathbf{Z}$ and $x \equiv y \mod n$. For any nonnegative integer k, let y^k represent the kth power of y as a member of $\mathbf{Z}_n^*$, and let x^k be the k-th power of x as an integer. Then $x^k \equiv y^k \mod n$. (Proof omitted.)* ∎

For example, $y^3 = ((y^2 \mod n)y \mod n)$. With $n = 6$ and $x = 10$, $y = 4$, $y^2 = 4$, and $y^3 = 4$. $x^3 = 1000$, and 6 divides $x^3 - y^3 = 996$.

Or, $x = 11$, $k = 4$, $n = 8$, so $y = 3$, $((((3 \times 3 \mod 8) \times 3) \mod 8) \times 3) \mod 8 = ((1 \times 3 \mod 8) \times 3) \mod 8 = 1$, whereas $11^4 = 14641$, and $14641 - 1 = 14640$, which is divisible by 8.

(12.10) Theorem *If p is prime and $x \in \mathbf{Z}$ is not divisible by p then $x^{p-1} \equiv 1 \mod p$.*

Proof. Let $y = x \mod p$. Since $p \not| x$, $gcd(p,x) = 1$ and $gcd(p,y) = 1$ (Lemma 12.7): $y \in \mathbf{Z}_p^*$.
By Corollary 12.5, $y^{p-1} = 1$ in $\mathbf{Z}_p^*$. By Lemma 12.9, $x^k \equiv 1 \mod p$. Q.E.D.

13 Normal subgroups and quotient groups

The relation $x \equiv y \mod n$ is preserved under the group operation (addition of integers) (Lemma 8.10).

What sort of subgroups H of a group G have a similar property, that the relation $x^{-1}y \in H$ is preserved under the group operation of G?

It would mean that for all $x, x', y, y' \in G$

$$(x^{-1}x' \in H \ \ \& \ \ y^{-1}y' \in H) \ \Rightarrow \ (xy)^{-1}(x'y') \in H. \tag{13.1}$$

(13.2) Definition *Given a group G, consider a mixed sequence of group elements and subsets of G, such as $a_1, a_2, A_3, a_4, A_5, A_6, a_7$, for eaample, where $a_i \in G$ and $A_j \subseteq G$. Then the* product set $a_1a_2A_3a_4A_5A_6a_7$ *is*

$$\{a_1a_2x_3a_4x_5x_6a_7 \colon x_3 \in A_3, x_5 \in A_5, x_6 \in A_6\}$$

The term 'product set' is not generally used in textbooks.
The following properties of product sets will help shorten the arguments.

[1] These notes make a bad job of arithmetic mod n

(13.3) Lemma (i) *Formation of product sets is an associative operation.*
(ii) *Formation of product sets is monotonic with respect to the 'subset' relation.*
(iii) *If $e \in A$ then $B \subseteq AB$, and if $e \in B$ then $A \subseteq AB$.*
(iv) *If H is a subgroup of G, then $HH = H$.*

(Proof omitted, but the first two are explained below.) ∎

Explanation of (i): For example, $(AB)C = A(BC)$, $x((yA)(zB)) = ((xy)A)(zB))$, $x(Ay) = (xA)y$, — or in short, ABC, $xyAzB$, and xAy mean the same thing no matter how they are parenthesised.

Explanation of (ii): For example, if $A \subseteq X$, then $AB \subseteq XB$, $xABy \subseteq xXBy$, and so on.

Equivalent to formula 13.2, we can say if $x' \in xH$ and $y' \in yH$ then $x'y' \in (xy)H$. In the notation of product sets, bearing in mind that their formation is an associative operation (Lemma 13.3), so $xHyH$ need not be parenthesised,

$$xHyH \subseteq xyH \tag{13.4}$$

for all $x, y \in G$.

In particular, with $y = x^{-1}$,

$$xHx^{-1}H \subseteq H$$

for all $x \in G$. Since $e \in H$, $xHx^{-1} \subseteq xHx^{-1}H$ (Lemma 13.3 (iii)), so

$$xHx^{-1} \subseteq H \tag{13.5}$$

Replacing x by x^{-1} in Formula 13.5, we get

$$x^{-1}Hx \subseteq H$$

for all $x \in G$. By Lemma 13.3 (ii), we can multiply on the left by x and on the right by x^{-1} to get (simplfying, by the associative property (i) in that Lemma)

$$H \subseteq xHx^{-1}$$

for all $x \in H$. Together with Formula 13.5, this implies

$$xHx^{-1} = H \tag{13.6}$$

for all $x \in G$. Multiply each side on the right by x and simplify, getting

$$xH = Hx \tag{13.7}$$

for all $x \in G$. From this it follows that for all $x, y \in G$,

$$xHyH = x(Hy)H = x(yH)H = xyHH = xyH$$

using Lemma 13.3 (i) and (iv). This implies Formula 13.4, which is equivalent to Formula 13.1. Summarising:

(13.8) Lemma *The formula 13.1, which says that the property '$x^{-1}y \in H$' is preserved under the group operation, is equivalent to any of the formulae 13.4, 13.5, 13.6 or 13.7, or the following:*

every left coset of H is a right coset and vice-versa. (*)

Proof. The first five formulae have already been proved equivalent.

Assume Formula 13.7, Let C be a left coset: choose $x \in C$; then $C = xH = Hx$, so C is also a right coset. Similarly, every right coset is a left coset. so (*) holds.

Suppose that (*) holds. For any $x \in G$, xH is a left coset, and also a right coset. But it contains x, and the only right coset containing x is Hx. Therefore $xH = Hx$, and Formula 13.7 holds.

Thus (*) is equivalent to Formula 13.7, so all the six formulae are equivalent. Q.E.D.

(13.9) Definition *A subgroup H of a group G is a* normal subgroup of G *if for every $x \in G$, $xHx^{-1} = H$.*

One writes $H \lhd G$ when H is a normal subgroup of G.

Note that this is a relation between H and G, not a property of H on its own. It is formula 13.6 which is the 'official' definition of a normal subgroup, although Lemma 13.8 supplies six equivalent ways of defining a normal subgroup.

Example. In S_3, $\langle(123)\rangle$ (which is, actually, A_3, the alternating group), is a normal subgroup but $\langle(12)\rangle$ is not. These featured in a homework exercise.

If G is abelian then *every* subgroup is a normal subgroup.

(13.10) Definition *Suppose $H \lhd G$. Then the* quotient group G/H *consists of the left (or equivalently, the right) cosets $\{xH : x \in G\}$, together with the operation*

$$[x][y] = [xy].$$

This is a well-defined operation on the left cosets $[x]$, since if $[x] = [x']$ and $[y] = [y']$ then $[xy] = [x'y']$.

(13.11) Lemma *If $H \lhd G$ then G/H is a group. If G is finite then $|G/H| = |G| \div |H|$.*

Proof. First, the operation is associative:
$[x]([y][z]) = [x][yz] = [xyz] = [(xy)z] = [xy][z] = ([x][y])[z]$.

Next, there exists an identity, namely, $[e]$:
$[x][e] = [xe] = [x] = [ex] = [e][x]$.

Last, $[x^{-1}] = [x]^{-1}$:
$[x][x^{-1}] = [xx^{-1}] = [e] = [x^{-1}x] = [x^{-1]}[x]$. Q.E.D.

Examples. It is easy to check that $A_4 \lhd S_4$. The quotient group S_4/A_4 has just two elements.

In an abelian group A, every subgroup H is a normal subgroup and A/H is always defined. For example, $\mathbf{Z}/\langle 15\rangle$ is defined — it is cyclic of order 15.

14 First isomorphism theorem for groups

Consider the following groups:

- Symmetries of a non-square rectangle
- The subgroup $\{e, (12), (34), (12)(34)\}$ of S_4.
- The subgroup $\{e, (12)(34), (13)(24), (14)(23)\}$ of S_4.
- The group $\{e, a, b, c\}$ with this Cayley Table:

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

- Symmetries of a non-square rhombus
- $\mathbf{Z}_2 \times \mathbf{Z}_2$ (So far, this is unexplained.)

Now formulate a definition of when two groups are

(14.1) Definition *Let $G, \bullet$ and $G', *$ be two groups. A* homomorphism *from $G, \bullet$ to $G'.*$ is a map $h\colon G \to G'$ such that for every $x, y \in G$,*

$$h(x \bullet y) = h(x) * h(y)$$

The kernel *of h is the set of elements of G mapped to the identity e' of G':*

$$\text{kernel}(h) = \{x \in G\colon h(x) = e'\}$$

(14.2) Lemma *If $h\colon G \to G'$ is a homomorphism, then* (i) *$h(e) = e'$ (h takes the identity of G to the identity of G'), and* (ii) *for any $x \in G$, $h(x^{-1}) = (h(x))^{-1}$.*

Proof. (i) Let $f = h(e)$. $ff = h(e)h(e) = h(ee) = h(e) = f$. Multiply on the left by f^{-1}, to get $f = e'$, as required.

(ii) Let $h(x) = x'$, so $x'h(x^{-1}) = h(x)h(x^{-1}) = h(xx^{-1}) = h(e) = e'$. Multiply on left by $(x')^{-1}$: $h(x^{-1}) = (x')^{-1}$. Q.E.D.

(14.3) Examples of homomorphisms.

- $\mathbf{Z} \to \mathbf{Z}_n$; $x \mapsto x \mod n$.
- $S_n \to \{1, -1\}$; $\sigma \mapsto 1$ if σ is an even permutation, -1 if σ is odd.
- Let G be a group, x an element of G. Then the map $\mathbf{Z} \to G$; $k \mapsto x^k$ is a homomorphism.
- If $H \lhd G$, then $h\colon G \to G/H$; $x \mapsto [x]$ (i.e., $x \mapsto xH$), is a homomorphism.

- From $\mathbf{Z}_2$ (under addition mod 2) to $\{1, -1\}$ (under multiplication: $0 \mapsto 1$, $1 \mapsto -1$.

(14.4) Lemma *If $h\colon G, \bullet \to G', *$ is a homomorphism,*[2] *then* (i) *its range or image,* range(h)*, is a subgroup of G', and*
(ii) *its kernel,* kernel(h)*, is a* normal *subgroup of G.*

Proof. We shall drop the $\bullet$ and $*$ signs. Let $S = \text{range}(h)$. R.T.P.: (a) $e' \in S$ — true, since $h(e) = e'$ (Lemma 14.2 (i)).

(b) S is closed under $*$. Given $x', y' \in S$, let $x' = h(x)$ and $y' = h(y)$, so $x'y' = h(x)h(y) = h(xy)$. Therefore $x'y' \in S$: S is closed under $*$.

(c) S is closed under inverse since $h(x^{-1}) = (h(x))^{-1}$. (Lemma 14.2 (ii)).

Thus S is a subgroup of G'.

Let $K = \text{kernel}(h)$. R.T.P. (a) $e \in K$: true since $h(e) = e'$ (Lemma 14.2 (i)).
(b) K is closed under '$\bullet$': given $x, y \in K$, $h(xy) = h(x)h(y) = e'e' = e'$, so $xy \in K$.
(c) K is closed under inverse: if $x \in K$, then $h(x) = e'$, and $h(x^{-1}) = (e')^{-1} = e'$ (Lemma 14.2 (ii)).

Thus K is a subgroup of G.
(d) $K \lhd G$: R.T.P. for any $x \in G$, $xKx^{-1} = K$.

For any $x \in G$, $k \in K$, $h(xkx^{-1}) = h(x)h(k)(h(x))^{-1} = h(x)e'(h(x))^{-1} = e'$, so $xkx^{-1} \in K$. Therefore

(e) $xKx^{-1} \subseteq K$ for any $x \in G$,

Replacing x by x^{-1}, we get $x^{-1}Kx \subseteq K$ for any $x \in G$. Given $k \in K$, it follows that $x^{-1}kx \in K$. Let k' denote $x^{-1}kx$, so $k = xk'x^{-1}$, i.e., $k \in xKx^{-1}$. This holds for all $k \in K$, so

(f) $K \subseteq xKx^{-1}$ for any $x \in G$.
(e) and (f) together imply (d), so $K \lhd G$. Q.E.D.

(14.5) Lemma *Let $h\colon G \to G'$ be a homomorphism, and let $H \lhd G$. Write π for the homomorphism, $G \to G/H$; $x \mapsto [x]$ (note $[x] = xH = Hx$).*

If $H \subseteq$ kernel(h) *then* (i) *there exists a unique homomorphism $\overline{h}\colon G/H \to G'$ satisfying $h = \overline{h} \circ \pi$, and (ii)* range$(h) =$ range$(\overline{h})$.

Proof. (i) The condition $h = \overline{h} \circ \pi$ means simply that $\overline{h}([x]) = h(x)$ for all $x \in G$, so $\overline{h}$ is uniuely defined if it exists. We must check that $\overline{h}([x])$ is well-defined, i.e., if $x^{-1}x' \in H$ then $h(x) = h(x')$.

Suppose $x^{-1}x' \in H$. Since $H \subseteq$ kernel(h), $h(x^{-1}x') = e'$, the identity of G'. Therefore $(h(x)^{-1})h(x') = (h(x))^{-1}h(x') = e'$, from Lemma 14.2 (ii), so $h(x) = h(x')$, as required: $\overline{h}$ is well-defined.

Next, $\overline{h}$ is a homomorphism: given $x, y \in G$, R.T.P. $\overline{h}([x][y]) = \overline{h}([x])\overline{h}([y])$.

$$\overline{h}([x][y]) = \overline{h}([xy]) = h(xy) = h(x)h(y) = \overline{h}([x])\overline{h}([y]),$$

as required.

(ii) $\quad \text{range}(\overline{h}) = \{\overline{h}([x])\colon x \in G\} = \{h(x)\colon x \in G\} = \text{range}(h).$ $\quad$ Q.E.D.

[2] In other words, $\bullet$ is the group operation in G and $*$ that in G'.

(14.6) Definition *An* isomorphism *between two groups G, G' is a bijective homomorphism from G onto G'.*

When an isomorphism exists, G and G' are called isomorphic, *written $G \cong G'$.*

(14.7) Corollary (First Isomorphism Theorem for groups). *If $h\colon G \to G'$ is a homomorphism, then* range$(h) \cong G/$kernel(h).

Proof. Let $H =$ kernel(h). In the notation of Lemma 14.5, there is a unique homorphism $\overline{h}\colon G/H \to G'$, satisfying $\overline{h}([x]) = h(x)$ for each $x \in G$.

For any $x, x' \in G$, if $\overline{h}([x]) = \overline{h}([x'])$ then $h(x) = h(x')$, so $(h(x))^{-1}h(x') = e'$, i.e. $h(x^{-1})h(x') = e'$ (Lemma 14.2 (ii)), so $h(x^{-1}x') = e'$; $x^{-1}x' \in$ kernel(h): $[x] = [x']$. Therefore $\overline{h}$ is one-to-one. Since range$(\overline{h}) =$ range(h), $\overline{h}$ is an isomorphism from $G/$kernel(h) onto range(h). Q.E.D.

(14.8) Lemma *Isomorphism of groups is an equivalence relation. (Proof easy, omitted.)*

One simple example of the First Isomorphism Theorem is where the homomorphism h is the identity map $G \to G$. This map is surjective, and its kernel is $\{e\}$, so

(14.9) Lemma $G \cong G/\{e\}$. ∎

Applications. Let us consider the examples given in paragraph 14.3

- $\mathbf{Z} \to \mathbf{Z}_n$; $x \mapsto x \mod n$. This is surjective, and its kernel is $\langle n \rangle$, so $\mathbf{Z}_n \cong \mathbf{Z}/\langle n \rangle$.
- $S_n \to \{1, -1\}$; $\sigma \mapsto 1$ if σ is an even permutation, -1 if σ is odd. The kernel of this map is A_n, and if $n \geq 2$ this is surjective, and $S_n/A_n \cong \{1, -1\}$.
- Let G be a group, x an element of G. Then the map $\mathbf{Z} \to G$; $k \mapsto x^k$ is a homomorphism. The range of this map is $\langle x \rangle$. If $|x| = n < \infty$ then the kernel is $\langle n \rangle$ and $\langle x \rangle \cong \mathbf{Z}/\langle n \rangle \cong \mathbf{Z}_n$ (from the first example, using transitivity of isomorphism).

  If $|x| = \infty$ then the kernel is $\langle 0 \rangle$, so $\langle x \rangle \cong \mathbf{Z}/\langle 0 \rangle \cong \mathbf{Z}$ (Lemma 14.9).
- If $H \lhd G$, then $h\colon G \to G/H$; $x \mapsto [x]$ (i.e., $x \mapsto xH$) is a homomorphism. Nothing much to say about this example.
- From $\mathbf{Z}_2$ (under addition mod 2) to $\{1, -1\}$ (under multiplication: $0 \mapsto 1$, $1 \mapsto -1$). This is obviously an isomorphism. Combining with the A_n examplw, $S_n/A_n \cong \mathbf{Z}_2$

15 Prime factorisation theorem

Recall Definition 12.6: a *prime number* is an integer $p \geq 2$ whose only positive factors are 1 and p.

(15.1) Lemma *Every integer $n > 1$ can be expressed as a product of primes.*

Proof. Uses *course-of-values induction*, which means that in order to prove that n is a product of primes, we can assume that every m, $2 \leq m < n$, is a product of primes.

Given $n \geq 2$, if n is prime then it is a product of primes. Otherwise, there exist k, m, $2 \leq k, m < m$, with $n = km$. Then by induction both k and m can be expressed as products of primes, so n can. Q.E.D.

(15.2) Lemma (i) *If a and b are nonzero integers and p is a prime and $p|ab$, then $p|a$ or $p|b$.* (ii) *If $n_1 \cdots n_k$ is a product of $k \geq 2$ integers, and p is a prime, and $p|n$, then $p|n_j$ for some j.*

Proof. (i) Suppose $p \not| a$, so $\gcd(p, b) \neq p$; then $\gcd(p, a)$ is another divisor of p, namely, 1. If $\gcd(p, b) = 1$ then $\gcd(p, ab) = 1$ (Lemma 11.5), whereas $p|ab$, so $\gcd(p, b) \neq 1$: the gcd is p, and $p|b$.

(ii) By induction on k. For the inductive step, suppose $p|n_1 n_2 \cdots n_k n_{k+1}$. If p divides one of the n_j, $j \leq k$, we are finished. Otherwise, by the inductive hypothesis, p does not divide the product $n_1 n_2 \cdots n_k$, call it a, whereas $p|an_{k+1}$, so by part (i) $p|n_{k+1}$. Q.E.D.

(15.3) Corollary *If p is a prime dividing a product $q_1 q_2 \ldots q_\ell$ of primes, then $p = q_j$ for some j.*

Proof. p divides one of the q_j. But q_j is prime, so its only factors are 1 and q_j, and $p \neq 1$, so $p = q_j$. Q.E.D.

(15.4) Theorem (the Prime Factorisation Theorem). *If $n \geq 2 \in \mathbf{Z}$ then n be expressed as a product $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where $p_1 < p_2 < \ldots p_k$ are primes, and e_j are positive integers, and the primes p_j and exponents e_j are unique.*

Proof. We already know that at least one such factorisation exists (Lemma 15.1). It remains to show that the factorisation is unique, which is proved by course-of-values induction on n.

If n is prime, then it has no factors except itself and 1, so $k = 1$, $e_1 = 1$, and $p_1 = n$; the factorisation is unique.

Suppose that n is not prime and has possibly different prime factorisations:

$$n = p_1^{e_1} \cdots p_k^{e_k} = q_1^{f_1} \cdots q_\ell^{f_\ell}. \tag{15.5}$$

Without loss of generality, $p_1 \leq q_1$. Since p_1 divides n, p_1 equals one of the q_j (Corollary 15.3), and if $j \geq 2$ then $p_1 < q_j$. Therefore $p_1 = q_1$.

Write $n = p_1^{e_1} m = p_1^{f_1} m'$. Since m and m' are products of primes not including p_1, p_1 doesn't divide m nor m'. If $e_1 < f_1$ then $m = p_1^{e_1 - f_1} m'$, so p_1 would divide m, and if $f_1 < e_1$ then p_1 would divide m' for similar reasons. Hence $e_1 = f_1$ and $m = m'$.

If $m = 1$ then $m' = 1$, $k = \ell = 1$, and n can be factorised only as $p_1^{e_1}$.. Otherwise, $m = m' > 1$, so:

$$m = p_2^{e_2} \cdots p_k^{e_k} = m' = q_2^{f_2} \cdots q_\ell^{f_\ell}.$$

Under the inductive hypothesis applied to m, $k-1 = \ell - 1$, $p_j = q_j$, and $e_j = f_j$ for $2 \le j \le k$. Therefore $k = \ell$, $p_j = q_j$, and $e_j = f_j$ for $1 \le j \le k$. The inductive step is proved. Q.E.D.

The following lemma is included here, because it is just about factorising numbers, though it will be applied to proving the existence of p-subgroups (Sylow theorem).

(15.6) Lemma *Suppose that n is a positive integer, p is a prime, and p^k is the highest power of p dividing n, where $k \ge 1$. Then*

$$p \not| \binom{n}{p^k}.$$

Proof. (This proof will be a standard calculation by cancelling out, not a formal exercise in number theory.) The binomial coefficient is

$$\binom{n}{p^k} = \frac{n(n-1)\cdots(n-p^k+1)}{p^k(p^k-1)\cdots 2\cdot 1}.$$

It is enough to show that the total power of p in the numerator equals that in the denominator. For $1 \le r \le p^k - 1$, we match the term r in the denominator with the term $n-r$ in the numerator, simply to show that they contain the same power of p:

Let $p^s, s \ge 0$ be the highest power of p dividing r, so $r = p^s t$, where p does not divide t, and $s < k$ since $r < p^k$. Also, $n - r = p^k m - p^s t = (p^{k-s}m - t)p^s$, so p^s divides $n-r$. But $(p^{k-s}m - t) \mod p = (-t) \mod p \neq 0$, so p does not divide $p^{k-s}m - t$, therefore p^{s+1} does not divide $n-r$. In other words, p^s is the highest power of p dividing $n-r$.

This leaves one term in the numerator and denominator respectively: n and p^k. The highest power of p dividing n is p^k. Hence all occurrences of p in the numerator are cancelled out by occurrences in the denominator. Q.E.D.

16 A Sylow theorem

(16.1) Group acting on a set. A *left action* of a group G on a set S is a homomorphism of G into the group of bijective maps from S onto S. Given $g \in G$, we write T_g for the corresponding map of S.

(16.2) Note that $T_e = \iota_S$, the identity map on S.

(16.3) Lemma *Given for every $g \in G$ a map $T_g \colon S \to S$, the correspondence $g \mapsto T_g$ is a left action if and only if* (a) *T_e is the identity map on S, and*

$$\text{(b)} \quad T_{hg}(s) = T_h(T_g(s))$$

for all $g, h \in G$, $s \in S$.

Proof. Let M be the group of bijective maps from S to S.

Only if: (a) T_e is the identity map on S because T is a homomorphism.

(b) For all $g, h \in G, s \in S$, $T_{hg}(s) = (T_h \circ T_g)(s) = T_h(T_g(s))$.

If: (b) is equivalent to

$$\text{(c)} \quad T_{hg} = T_h \circ T_g$$

for all $g, h \in G$. Therefore, for any $g \in G$,

$$T_g \circ T_{g^{-1}} = T_e = \iota_S = T_{g^{-1}} \circ T_g,$$

so every T_g is a bijective map, i.e., $T_g \in M$ for each $g \in G$, so T is a map from G into M.

So $T\colon G \to M$ is a map satisfying (c), i.e., a homomorphism. T is a left action. Q.E.D.

Examples If S is just the group itself, then we can define $T_g(s) = gs$, multiplication by g on the left, not a very interesting example.

A slightly more interesting example would be $T_g(s) = gsg^{-1}$, conjugation by g.

If $G = S_n$ and $S = \{1, \ldots, n\}$, we can define $T_\sigma(i) = \sigma(i)$, not very interesting.

If $G = S_n$ and S is the set of 2-element subsets of $\{1, \ldots, n\}$, we can define $T_\sigma(\{i, j\}) = \{\sigma(i), \sigma(j)\}$, not very interesting.

(16.4) Definition *Given $s \in S$, the* stabiliser G_s *of s is the set*

$$G_s = \{g \in G : T_g(s) = s\}$$

(16.5) Lemma *G_s is a subgroup of G.*

Proof. Note $e \in G_s$ since $T_e(s) = s$ always (paragraph 16.2). If $T_g(s) = s$ and $T_h(s) = s$ then $T_{hg}(s) = T_h(T_g(s)) = T_h(s) = s$ and $T_{g^{-1}}(s) = T_{g^{-1}}(T_g(s)) = T_e(s) = s$. Thus G_s is a subgroup of G. Q.E.D.

(16.6) Definition *The* orbit *of s is the set $\{T_g(s)\colon g \in G\}$.*

(16.7) Lemma *The orbits form a partition of S.*

Sketch of proof. The relation: $y = T_g(x)$ for some $g \in G$ is an equivalence relation on S, and its equivalence classes are the orbits of S. ∎

(16.8) Lemma *Given $s \in S$, $T_g(s) = T_h(s)$ if and only if the left cosets gG_s and hG_s are the same.*

Proof. If $gG_s = hG_s$, then $h = gx$ for some $x \in G_s$. Then $T_h(s) = T_{gx}(s) = T_g(T_x(s)) = T_g(s)$.

If $T_h(s) = T_g(s)$, then $T_{g^{-1}}(T_h(s)) = s$, so $T_{g^{-1}h}(s) = s$: $g^{-1}h \in G_s$, so $h \in gG_s$, so $gG_s = hG_s$. Q.E.D.

(16.9) Corollary *For any $s \in S$, $|\mathit{orbit}(s)| = |G|/|G_s|$.*

Proof. According to the above, the elements of the orbit of s correspond bijectively to the left cosets of G_s. Q.E.D.

In order to prove the Sylow Theorem below, we consider a set S consisting of all p^k-element subsets of G:

$$S = \{s \subseteq G \colon |s| = p^k\}$$

and the following left action on S:

$$T_g(s) = \{gx \colon x \in s\}.$$

(16.10) Lemma *This is a left action on S.*

Proof. Given $s \in S$, $g \in G$, for every $x \neq y \in s$, $gx \neq gy$. Therefore $|T_g(s)| = |s|$, so T_g is a map from S to S.

$T_e(s) = \{x \colon x \in s\} = s$, for all $s \in S$.

Given $h, g \in G$, $s \in S$,

$$T_g(s) = \{gx \colon x \in s\}, T_h(T_g(s)) = \{hy \colon y \in T_g(s)\} = \{h(gx) \colon x \in s\} = \{(hg)x \colon x \in s\} = T_{hg}(s).$$

By Lemma 16.3, T_g is a left action. Q.E.D.

There is one other fact about this particular left action:

(16.11) Lemma *With this particular left acion, for any $s \in S$, $|G_s| \leq |s|$.*

Proof. For any $g \in G_s$, $\{gx \colon x \in s\} = s$. Fix $a \in s$, so for any $g \in G_s$, $ga \in s$. Also, $|G_s| = |\{ga \colon g \in G_s\}|$, since $g \neq h \Rightarrow ga \neq ha$, Since $\{ga \colon g \in G_s\} \subseteq s$, it follows that $|G_s| \leq |s|$. Q.E.D.

(16.12) Theorem (a Sylow theorem). *Let G be a finite group, and let p be a prime divisor of $|G|$: suppose that p^k is the highest power of p dividing $|G|$. Then G has a subgroup of order p^k.*

Proof. We consider the set S and left action T_g introduced above. The cardinality of S is

$$|S| = \binom{n}{p^k}$$

and according to Lemma 15.6, this is not divisible by p. The orbits form a partition of S, so there exists an s the length of whose orbit is not divisible by p.

This implies that p does not divide $|G|/|G_s|$, so $|G_s|$ is divisible by p^k.

But according to Lemma 16.11, $|G_s| \leq p^k$. Therefore $|G_s| = p^k$, and G_s is the required subgroup. Q.E.D.

A concrete example. According to this theorem, S_8 contains a subgroup of order 128, by considering all 128-element subsets of S_8. This is a vast number of sets to consider, and the only reasonable example is to prove by this method that S_3 contains a subgroup of order 2.

S_3 can be generated by $a = (123)$ and $b = (12)$. These elements satisfy the relations $a^3 = e$, $b^2 = e$. We know immediately that $\{e, b\}$ is a subgroup of order 2, but we ignore that.

Also, $ab = (13) = ba^2$ and $a^2b = (23) = ba$.

$\{e, a, a^2\}$ is a subgroup of order 3, and the right coset containing b is $\{b, ab, a^2b\}$. This must be the other coset, since b has order 2 whereas e, a, a^2 have orders $1, 3, 3$, respectively. We can label the six group elements as e, a, a^2, b, ab, a^2b. The Cayley Table is easily constructed using the above four equations.

	e	a	a^2	b	ab	a^2b
e	e	a	a^2	b	ab	a^2b
a	a	a^2	e	ab	a^2b	b
a^2	a^2	e	a	a^2b	b	ab
b	b	a^2b	ab	e	a^2	a
ab	ab	b	a^2b	a	e	a^2
a^2b	a^2b	ab	b	a^2	a	e

The set S consists of all 2-element subsets of S_3. There are 15 of these, namely,

$$\{e, a\}, \{e, a^2\}.\{e, b\}, \{e, ab\}, \{e, a^2b\},$$
$$\{a, a^2\}, \{a, b\}, \{a, ab\}, \{a, a^2b\},$$
$$\{a^2, b\}, \{a^2, ab\}, \{a^2, a^2b\},$$
$$\{b, ab\}, \{b, a^2b\},$$
$$\{ab, a^2b\}.$$

The left action of S_3 consists of multiplying each of these sets on the left by the six elements of S_3. Note that the orbit lengths always divide the order of the group, so the orbit lengths can be $1, 2, 3$, or 6. Starting with $\{e, a\}$, and multiplying on the left by group elements, we get

$$\{e, a\}, \{a, a^2\}, \{a^2, e\}, \{b, a^2b\}, \{ab, b\}, \{a^2b, ab\}$$

Start again at $\{e, b\}$:

$$\{e, b\}, \{a, ab\}, \{a^2, a^2b\}$$

The next set in the orbit would be $b\{e, b\} = \{b, e\}$, so the orbit has at least 3 elements and fewer than 6: so this is the full orbit. Start again with $\{e, ab\}$.

$$\{e, ab\}, \{a, a^2b\}, \{a^2, b\}.$$

The next set is $\{b, a^2\}$, which has already been met.

Start again at $\{e, a^2b\}$.

$$\{e, a^2b\}, \{a, b\}, \{a^2, ab\}.$$

We have generated assl 15 sets in S, so we are finished. There is one orbit of length 6, and three odd-length orbits of length 3 each. According to the proof of Sylow's theorem, any one of the 9 sets in the odd-length orbits has stabiliser group of order 2. For example, let us calculate G_s where $s = \{a, a^2b\}$.

We need to find all g such that

$ga = a, g(a^2b) = a^2b$, or $ga = a^2b, ga^2b = a$.

The first pair of equations give $g = e$. For the second pair, $g = a^2ba^{-1} = a^2ba^2 = baa^2 = b$. Check that $ga^2b = a$: $ba^2b = bba = a$. Thus the stabiliser group for $\{a, a^2b\}$ is $\{e, b\}$.

Take one more set, such as $\{a, ab\}$. Solve $ga = ab$: $g = aba^{-1} = aba^2 = aab = a^2b$. Check $a^2bab = a^2b^2a^2 = a$. This gives us the subgroup $\{e, a^2b\} = \{e, (23)\}$.

17 Classification of finite abelian groups

It is the aim of this section to prove that every finite abelian group A can be expressed as a product of cyclic p_j-groups, where p_j are the prime divisors of $|A|$, and apart from the order of factors the product is unique.

The proof is too long to be covered fully in lectures. The full proof would be in four parts:

- Every nontrivial finite abelian group A is isomorphic to a product of p_j-groups, one for each prime divisor of $|A|$. (This is covered in lectures.)
- Every nontrivial finite abelian p-group is isomorphic to a product of cyclic p-groups. (The only part covered in lecturess is where every element has order 1 or p.)
- Given a product of abelian p_i-groups, and another isomorphic product of abelian q_j-groups, where the p_i and q_j are distinct then after reordering if necessary the corresponding factors are isomorphic.
- Given two finite products of cyclic p-groups, if they are isomorphic then after reordering if necessary the corresponding factors are isomorphic.

(17.1) Definition *Let p be a prime. A p-*group *G is a group such that for every $x \in G$, $|x|$ is a power of p.*

(17.2) Lemma *A finite group G is a p-group if and only if $|G|$ is a power of p.*

Proof. If: for every $x \in G$, $|x|$ divides $|G|$, so $|x|$ is a power of p.

Only if: suppose that $|G|$ is not a power of p, so there is another prime q dividing $|G|$. By the Sylow theorem, there is a subgroup H of G whose order is a positive power of q. Then for any $x \neq e \in H$, $|x|$ is a nonzero power of q, so G is not a p-group. Q.E.D.

Let A be a finite abelian group, $|A| = n$. Let $n = p_1^{e_1} \cdots p_k^{e_k}$, the prime factorisation.

17.1 A finite abelian group is a product of p_j-groups

(17.3) Lemma *For $1 \leq j \leq k$, let $n_j = n/p_j^{e_j}$. Then there exist integers $r_j, 1 \leq j \leq k$ such that*

$$r_1 n_1 + \ldots + r_k n_k = 1.$$

Proof. The numbers n_j have greatest common divisor 1, because any prime p dividing all of them also divides n, hence must be one of the p_j (Lemma 15.2): but p_j does not divide n_j.

The gcd of these numbers generates the subgroup $\langle n_1, \ldots, n_k \rangle$ of $\mathbf{Z}$, so $1 \in \langle n_1, \ldots, n_k \rangle$, but $\langle n_1, \ldots, n_k \rangle$ is the set of all expressions $r_1 n_1 + \ldots + r_k n_k$, $r_j \in \mathbf{Z}$. Q.E.D.

(17.4) Lemma *For $1 \leq j \leq k$, let $A_j = n_j A$, i.e., $\{n_j x \colon x \in A\}$.*

Each A_j is subgroup of A, a p_j-group, and its order divides $p_j^{e_j}$.

Proof. For any $x \in A_j$, $x = n_j y$ for some $y \in A$. Then $p_j^{e_j} x = ny = 0$, so $|x|$ divides $p_j^{e_j}$. Thus A_j is a p_j-group and $|A_j|$ is a power of p_j (Lemma 17.2). Since $|A_j|$ divides n by Lagrange's Theorem, it divides $p_j^{e_j}$.[3] Q.E.D.

The cartesian product $A_1 \times \cdots \times A_k$ is the set of all k-tuples $\{(x_1, \ldots, x_k) \colon x_j \in A_j\}$, an abelian group under componentwise addition.

(17.5) Lemma *A is isomorphic to the cartesian product $A_1 \times \cdots \times A_k$, and each A_j is an abelian p_j-group of order $p_j^{e_j}$.*

Proof. With A, n, n_j, r_j, and A_j as introduced above, consider two maps

$$\theta \colon A \to A_1 \times \cdots \times A_k; \quad x \mapsto (n_1 x, \ldots, n_k x)$$

and

$$\phi \colon A_1 \times \cdots \times A_k \to A; \quad (x_1, \ldots, x_k) \mapsto r_1 x_1 + \ldots + r_k x_k.$$

Both θ and ϕ are easily seen to be homomorphism.

From the definition of the numbers r_j, $\phi \circ \theta$ is the identity map on A. Therefore θ is injective, so $n = |A| \leq |A_1 \times \cdots \times A_k|$.

Now, for each j, $|A_j|$ divides $p_j^{e_j}$, and $|A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k|$ which divides n. Thus $n \leq |A_1| \cdots |A_k| \leq n$, so the product of the orders equals n.

Thus θ is an injective homomorphism from A into another group of order n; so θ is an isomorphism.

Finally, each $|A_j|$ divides $p_j^{e_j}$. If any $|A_j|$ were a proper divisor then $|A_1| \cdots |A_k|$ would be a proper divisor of n. So $|A_j| = p_j^{e_j}$ for each j. Q.E.D.

17.2 A finite abelian p-group is a product of cyclic groups

For this subsection A is a nontrivial finite p-group.

(17.6) Definition *In any aaelean group, a list $a_1, ldots, a_k$ of elements is* independent *if for all tuples $\alpha_1, \ldots, \alpha_k \in \mathbf{Z}$, if $\alpha_1 a_1 + \ldots + \alpha_k a_k = 0$, then $\alpha_1 a_1 = \ldots = \alpha_k a_k = 0$.*

In the lemme below, it is not necessary tht A be a p-group.

(17.7) Lemma *Given an abelian group A, if there exists an independent set of generators for A, then A is isomorphic to a product of cyclic groups.*

Proof. We consider the following homomorphism

$$\theta \colon \langle a_1 \rangle \times \cdots \times \langle a_k \rangle \to A; \quad (\alpha_1 a_1, ldots, \alpha_k a_k) \mapsto \alpha_1 a_1 + ldots + \alpha_k a_k.$$

Since the a_j are generators, θ is surjective. The kernel of θ is the set of tuples $(\alpha_1 a_1, \ldots, \alpha_k a_k)$ such that $\alpha_1 a_1 + \ldots + \alpha_k a_k = 0$. Since the a_j are independent, the kernel is $\{(0, \ldots, 0)$, so θ is injective. Therefore θ is an isomorpism, and the groups $\langle a_j \rangle$ are cyclic. Q.E.D.

So the problem is to locate a set of independent generators for A. One case is relatively easy:

[3] To prove this, write $k = |A_j|$, $q = p_j^{e_j}$ and $n = qm$, so $\gcd(k, m) = 1$. Therefore $kr + ms = 1$ for some $r, s \in \mathbf{Z}$, and k divides qm, so it divides $qkr + qms = q$, i.e. k divides q.

(17.8) Lemma *Suppose that A is a finite abelian group in which every element except 0 has order p, so $|A| = p^r$ for some r. Then A is isomorphic to $\mathbf{Z}_p^r = \mathbf{Z}_p \times \cdots \times \mathbf{Z}_p$ (r factors),*

Proof. Start with some $a_1 \neq 0$ in A. If $A = \langle a_1 \rangle$, stop. Otherwise choose some $a_2 \in A \backslash \langle a_1 \rangle$. If $A = \langle a_1, a_2 \rangle$ stop, otherwise choose some $a_3 \in A \backslash \langle a_1, a_2 \rangle$.

The general step is, having chosen $a_1, \ldots, a_j$ in this way, if these do not together generate A then choose some $a_{j+1} \in A \backslash \langle a_1, \ldots, a_j \rangle$. Since all these a_j are distinct and A is finite, this process must terminate having chosen a set $\langle a_1, \ldots, a_k \rangle$ of generators for A. Each a_j has order p. R.T.P: they are independent.

Othereise there exist α_j, $0 \leq \alpha_j \leq p-1$ for each j and $\alpha_1 a_1 + \ldots + \alpha_k a_k = 0$. Choose the largest i such that $\alpha_i \neq 0$. If $i = 1$ then $\alpha_1 a_1 = 0$ so $\alpha_1 = 0$ and all the α_j are 0.

Otherwise, since α_i has a multiplicative inverse modulo p,[4] there exists a β such that $\beta \alpha_i \equiv 1 \mod p$. Multiplying across by β,

$$\beta \alpha_1 a_1 + \ldots \beta \alpha_{i-1} + a_i = 0.$$

It follows that $a_i \in \langle a_1, \ldots, a_{i-1} \rangle$, which is false. Therefore the kernel is trivial and θ is an isomorphism.

Finally, since the cartesian product has cardinality p^k, $k = r$. Q.E.D.

The remainder of this section is somewhat more difficult. To prove the general result, that is, a finite abelian p-group A is a product of cyclic p-groups, we use course-of-values induction on $|A|$. The induction involves considering the subgroup $pA = \{px \colon x \in A\}$.

(17.9) Lemma *If A is a finite nontrivial abelian p-group then pA is a proper subgroup of A.*

Proof. The map $x \mapsto px$ is a homomorphism $a \to A$. Since A is finite, to show that its range is a proper subgroup of A it is enough to show that the map is not injective. Choose any $x \neq 0$ in A. Let $|x| = p^{k+1}$ and $y = p^k x$. Then $y \neq 0$ but $py = 0$, so the map is not injective. Q.E.D.

If pA is the trivial group, then every element of A has order 1 or p and we already know the structure of A. Otherwise, we assume by induction that pA admits an independent set of generators, that is an independent set of elements

$$pa_1, \ldots, pa_k$$

The next step is to prove that the elements $a_1, \ldots, a_k$ of A are independent elements of A. The proof (whoci is omitted) boils down to arguing that if these elements are int independent, then one of the generators for pA, pa_1, say, could be replaced by one of lower order. This is impossible on cardinality grounds.

Next we would consider the subgroup B of A generated by the elements a_j. If $B = A$ then we are finished. Otherwise A/B is a group of the kind considered in Lemma 17.8. We can use the construction of that lemma to extend the generators a_j to a list of independent generators for A.

[4] Recall that $\mathbf{Z}_p^*$ consists of all numbers between 1 and p.

17.3 First part of the uniqueness result.

We now know the 'existence' part of the classification theorem. We need to know the 'uniqueness' part.

(17.10) Lemma *Suppose A is a nontrivial finite abelian group, and $|A| = n = p_1^{e_1} \cdots p_k^{e_k}$. We know that A is isomorphic to one product $A_1 \times \cdots \times A_k$ of p_j groups. Suppose that it is isomorphic to another product $A'_1 \times A'_\ell$ of nontrivial q_i-groups. where the primes q_i are distinct. Then $k = \ell$ and after reordering if necessary, $A_j \cong A'_j$ for $1 \leq j \leq k$.*

Proof. Let θ be an isomorphism. If $x \neq 0 \in A_j$, then $|x|$ is a power of p_j, the same goes for $\theta(x)$, which belongs to one of the groups A'_i, which is also a p_j-group. Thus $q_i = p_j$ and θ maps A_j into A'_i. Since $|A'_i|$ divides n, $|A'_i| \leq p_j^{e_j}$, Also $|A_j| \leq |A'_i|$. Therefore $|A_j| = |A_i|$ and θ carries A_j isomorphically onto A'_i.

Thus for each j there exists a unique i such that A_j is isomorphic to A'_i. There can be no A'_i to which no A_j is isomorphic since otherwise $|A'_1 \times \cdots \times A'_\ell| > n$. Thus $k = \ell$ and after reordering the A'_i if necessary $A_j \cong A'_j$ for $1 \leq j \leq k$. Q.E.D.

17.4 The final part of the uniqueness result

To conclude the result, we need to prove the following:

(17.11) Lemma *Given two finite abelian p-groups, expressed as products of cyclic groups,*

$$\mathbf{Z}_{p^{r_1}} \times \cdots \times \mathbf{Z}_{p^{r_k}} \tag{17.12}$$

and

$$\mathbf{Z}_{p^{s_1}} \times \cdots \times \mathbf{Z}_{p^{s_\ell}} \tag{17.13}$$

where $r_1 \geq r_2 \geq \ldots \geq r_k \geq 1$ and $s_1 \geq s_2 \ldots \geq s_\ell \geq 1$, then $k = \ell$ and for $1 \leq j \leq k$, $r_j = s_j$.

First

(17.14) Lemma *Let G_i, $1 \leq i \leq k$ be groups, and for each i let H_i be a normal subgroup of G_i. Then $H_1 \times \cdots \times H_k \lhd G_1 \times \cdots \times G_k$ and*

$$(G_1/H_1) \times \cdots \times (G_k/H_k) \cong (G_1 \times \cdots \times G_k)/(H_1 \times \cdots \times H_k).$$

(Proof: can be deduced from a homework exercise.)

(17.15) Lemma *Let*

$$A = \mathbf{Z}_{p^{r_1}} \times \cdots \times \mathbf{Z}_{p^{r_k}}$$

For any $r \geq 1$, $(p^{r-1}A)/(p^r A)$ is isomorphic to $\mathbf{Z}_p^t$, where t is the number of exponents r_j such that $r_j \geq r$.

Proof. By the above lemma $(p^r A)/(p^{r+1}A)$ is isomorphic to the product of the following quotient groups:

$$(p^{r-1}\mathbf{Z}_{p^{r_j}})/(p^r \mathbf{Z}_{p^{r_j}})$$

If $r \leq r_j$, then $p^{r-1}\mathbf{Z}_{p^{r_j}}$ is cyclic of order $r_j + 1 - r$, and the other factor group is cyclic of order $r_j - r$. Hence the quotient group has order p, and is isomorphic to $\mathbf{Z}_p$.

If $r > r_j$, then both groups are trivial and the quotient is trivial. Hence $(p^{r-1}A)/(p^r A)$ is isomorphic to as many copies of $\mathbf{Z}_p$ as there are $r_j \geq r$. Q.E.D.

(17.16) Definition *$d(r, A)$ is the number of copies of $\mathbf{Z}_p$ occurring in $(p^{r-1}A)/(p^r A)$. (Equivalently, the t such that $(p^{r-1}A)/(p^r A)| = p^t$.)*

Proof of Lemma 17.11. Referring to the two possibly different decompositions of A in equations 17.12 and 17.13 above,

We want to prove that the two non-increasing sequences

$$r_1, r_2, \ldots, r_k$$

and

$$s_1, s_2, \ldots, s_\ell$$

are identical.

Alternatively, let A denote the product of the $\mathbf{Z}_{p^{r_j}}$, and A' the product of the $\mathbf{Z}_{p^{s_j}}$, and suppose the sequences differ: R.T.P. A and A' are not isomorphic.

Choose $i \geq 1$ maximal subject to $p_j = q_j$, $1 \leq j < i$. Then either

(a) $i = \ell + 1 \leq k$, (b) $i = k + 1 \leq ell$, (c) $i \leq k, \ell$ but $r_i > s_i$, or (d) $i \leq k, \ell$ but $s_i > r_i$.

(a) $d(A, r_i) \geq i$ but $d(A', r_i) = \ell$. (b) Similar.

(c) $d(A, r_i) \geq i$, but $d(A', r_i) < i$. (d) Similar. Q.E.D.

18 Rings

(18.1) Definition *A* ring $R, +, *$ *is a nonempty set with two operations, addition and multiplication, such that $R, +$ is an abelian group and $R, *$ is a semigroup. As with groups, the multiplication sign is often omitted, so xy means $x * y$.. The additive identity is denoted* 0.

Distributive laws *connect these two operations:*

$$x(y + z) = (xy) + (xz)$$

$$(x + y)z = (xz) + (yz)$$

.

Examples.

- $\mathbf{Z}, +, *$, the ring of integers.
- $2\mathbf{Z}$, the ring of even integers.
- $\mathbf{Z}_n, +, *$, with addition and multiplication reduced modulo n.
- $\mathbf{Z}[\sqrt{2}] = \{a + b\sqrt{2} \colon a, b \in \mathbf{Z}\}$, with $(a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2}$, and $(a + b\sqrt{2})(c + d\sqrt{2}) = (ab + 2cd) + (bc + ad)\sqrt{2}$,
- Let S be a nonempty set. $M(S, \mathbf{Z})$, the set of all mappings from S to $\mathbf{Z}$ with pointwise addition and multiplication, i.e., $(f + g)(x) = f(x) + g(x)$ and $fg(x) = f(x)g(x)$ for all $x \in S$.
- $M(2, \mathbf{R})$, the ring of 2×2 matrices with real-valued entries.
- The sets $\mathbf{Q}, \mathbf{R}, \mathbf{H}$ mentioned at the beginning of this course.

Recall the rules for addition and multiplication of 2×2 matrices.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}, \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

(18.2) Definition *A* unity *in a ring $R, +, *$ is a two-sided multiplicative identity.*

Not every ring possesses a unity. For example, the ring of even integers has no unity. We already know from results about semigroups that the unity must be unique.

(18.3) Definition *A* subring *of a ring $R, +, *$ is a nonempty subset S which is closed under addition, negation, and multiplication.*

Remark. Every subring of R contains 0 and is a ring under the operations of R.
In the ring $M(2, \mathbf{R})$, the following is a subring:

$$\left\{ \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} : a, b \in \mathbf{R} \right\}$$

It has infinitely many left identities:

$$\begin{bmatrix} 1 & z \\ 0 & 0 \end{bmatrix}$$

But, of course, no right identities and no unity.

(18.4) Notation *A* unital *ring is a ring with a unity.*

More examples of subrings: $\mathbf{Z}$ is a subring of $\mathbf{Q}$ which is a subring of $\mathbf{R}$ which is a subring of $\mathbf{C}$ which is a subring of $\mathbf{H}$. $\mathbf{Z}[\sqrt{2}]$ is a subring of $\mathbf{R}$ and of $\mathbf{C}$.
The following simple facts hold for a ring R.

- The zero element is unique, because it is the identity in the group $R, +$

- Every $a \in R$ has a unique negative $-a$.
- Addition is cancellative.
- Write $x - y$ for $x + (-y)$.
- For any $a, b \in R$, the equations $a+x = b$ and $x+a = b$ both have the unique solution $x = b-a$.
- For any integers m, n, $(m+n)a = (ma)+(na)$, $m(a+b) = (ma)+(mb)$, and $m(na) = (mn)a$.

All of these facts hold for any abelian group $R, +$. The following are specific to rings.

(18.5) Lemma *In a ring R,*

(a) $0x = x0 = 0$
(b) $x(-y) = (-x)y = -(xy)$.
(c) $(-x)(-y) = xy$
(d) $x(y - z) = (xy) - (xz)$ *and* $(x - y)z = (xz) - (yz)$.

Proof. (a) $(0x)+(0x) = (0+0)x = 0x$: therefore $0x = (0x)+(0x)+-(0x) = 0x+-(0x) = 0$. Similarly, $x0 = 0$.
(b) $0 = x0 = x(y + -y) = xy + x(-y)$. Therefore $x(-y) = -(xy)$. Similarly, $(-x(y - (xy)$.
(c) $(-x)(-y) = -(x(-y)) = -(-(xy)) = xy$.
(d) $x(y - z) = x(y + (-z)) = xy + x(-z) = xy + -(xz) = xy - xz$.
Similarly, $(x - y)z = xz - yz$. Q.E.D.

19 Zero divisors, integral domains, and fields.

It is possible for a ring R to contain two elements a and b, neither of them zero such that $ab = 0$.

Examples. In $\mathbf{Z}_4$, $2 * 2 = 0$.
In $M(2, \mathbf{R})$,

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

Correction. What have to date been called cartesian products of groups should have been called direct products of groups.

(19.1) Definition *The* direct product of rings *Given two rings $R_1, +_1, *_1$ and $R_2, +_2, *_2$, their* direct product *is the cartesian product $R_1 \times R_2$ together with the operations $(x_1, x_2) + (y_1, y_2) = (x_1 +_1 y_1, x_2 +_2 y_2)$, and $(x_1, x_2) * (y_1, y_2) = (x_1 *_2 y_1, x_2 *_2 y_2)$.*

(19.2) Lemma *The direct product of two unital rings is a unital ring (Trivial).* ∎

Direct product rings have many zero divisors, since $(a, 0) * (0, b) = (0, 0)$ for all a, b.

(19.3) Definition *An* integral domain *is a ring which* (a) *is commutative,* (b) *is unital (i.e., a ring with unity), and* (c) *has no zero divisors.*

Examples.

- $\mathbf{Z}, \mathbf{Q}, \mathbf{R}$ are integral domains,
- $\mathbf{Z}[\sqrt{2}]$ is an integral domain.
- $2\mathbf{Z}$ is not because it is not unital.
- $\mathbf{H}$ is not because it is not commutative.
- $M(2, \mathbf{R})$ is not because it is not commutative and has zero divisors.
- $\mathbf{Z}_4$ is not because $2^2 = 0$ in $\mathbf{Z}_4$.
- $\mathbf{Z}_2$, and in general $\mathbf{Z}_p$ for any prime p, is an integral domain.

(19.4) Lemma *A commutative unital ring D is an integral domain if and only if it satisfies the following* cancellation law: *for any $a, b, c \in D$, if $ab = ac$ (or $ba = ca$) and $a \neq 0$, then $b = c$.*

Proof. If: given $ab = 0$, if $a \neq 0$ then $ab = a0$ so $b = 0$, and if $b \neq 0$ then $ab = 0b$ so $a = 0$. Therefore D has no zero divisors.

Only if: given $ab = ac$, $ab - ac = 0 = a(b - c)$. Since a is not a zero divisor, $b - c = 0$, so $b = c$. Q.E.D.

(19.5) Definition *A* field *is a commutative ring F in which the subset $F\backslash\{0\}$ forms a group under multiplication.*

In other words, for every $x \in F$, if $x \neq 0$ then x has a multiplicative inverse. Of course, since $0x = 0$ for all x, 0 does not have a multiplicative inverse (unless the ring is trivial, which we ignore).

(19.6) Lemma *Every field F is a nontrivial unital ring without zero divisors, i.e., an integral domain.*

Proof. $F\backslash\{0\}$ contains a multiplicative identity, that is, a unity. Given $x, y \neq 0$, $xy \neq 0$ since $F\backslash\{0\}$ is closed under multiplication. Q.E.D.

(19.7) Lemma *Let F be a commutative unital (nontrivial) ring in which every nonzero element has a multiplicative inverse. Then F is a field.*

Proof. R.T.P. $F\backslash\{0\}$ is closed under multiplication. Given x and y with $x \neq 0$, if $xy = 0$ then $x^{-1}xy = y = 0$. Hence if $x, y \neq 0$ then $xy \neq 0$.

Examples.

- $\mathbf{Z}, \mathbf{Z}[\sqrt{2}], 2\mathbf{Z}$, and $\mathbf{Z}_4$ are not fields.
- $\mathbf{Q}$, $\mathbf{Q}[\sqrt{2}]$, $\mathbf{R}$, and $\mathbf{C}$ are fields.
- $\mathbf{H}$ is not a field, because, although $\mathbf{H}\backslash\{0\}$ is a group under multiplication, it is not commutative.
- For any prime p, $\mathbf{Z}_p$ is a field.

To show that $\mathbf{Q}[\sqrt{2}]$ is a field: First, that $\sqrt{2}$ is irrational. This result is due to Euclid.

Note. This proof differs from the clumsy proof given in class.

(19.8) Lemma *There exist no integers p and q such that $(p/q)^2 = 2$.*

Proof. Equivalently, suppose that p and q are nonzero integers; R.T.P. $p^2 \neq 2q^2$. Write $p = 2^r s$ and $q = 2^t u$ where $r, t \geq 0$ and s, u are odd.

So we must show that

$$2^{2r}s^2 \neq 2^{2t+1}u^2$$

where s and u are odd integers. If $r \leq t$, then

$$s^2 \neq 2^{2t-2r+1}u^2,$$

since the left-hand side is odd and the right-hand side is even. Therefore, since $\mathbf{Z}$ is cancellative, $2^{2r}s^2 \neq 2^{2t+1}u^2$.

Otherwise, $r > t$, and

$$2^{2r-2t-1}s^2 \neq u^2,$$

since the left-hand side is even and the right-hand side is odd. Therefore, since $\mathbf{Z}$ is cancellative, $2^{2r}s^2 \neq 2^{2t+1}u^2$. Q.E.D.

Now to show that $\mathbf{Q}[\sqrt{2}]$ is a field, i.e., every nonzero element has a multiplicative inverse.

Given $a + b\sqrt{2}]$ where a and b are not both zero, its multiplicative inverse is $(a - \sqrt{2})/(a^2 - 2b^2)$. The denominator is nonzero because 2 is irrational, and the product is $(a^2 - 2b^2)/(a^2 - 2b^2) = 1$.

(19.9) Lemma *Every (nontrivial) finite integral domain D is a field.*

Proof. Since D is commutative and unital with unity, call it 1, different from 0, it is enough to show that every nonzero a has a multiplicative inverse.

Since $a \neq 0$ and D is cancellative, the map $D \to D; x \mapsto ax$ is injective. Since D is finite, the map is bijective, so there exists a y such that $xy = 1$. Q.E.D.

(19.10) Lemma $\mathbf{Z}_n$ *is an integral domain (and a field) if and only if n is prime.*

Proof. Either $n = 1$, n is composite, or n is prime. Since $\mathbf{Z}_1 \backslash \{0\}$ is empty, $\mathbf{Z}_1$ is not a field. If $n > 1$ is composite, there exist k, m, $2 \leq k, m < n$, so $km = n$. Then k and m are zero divisors in $\mathbf{Z}_n$.

Suppose that n is prime and $0 \leq k, m < n$ and $km = 0$ in $\mathbf{Z}_n$. R.T.P.: $k = 0$ or $m = 0$. Since $km = 0$ in $\mathbf{Z}_n$, n divides km. By Lemma 15.2, n divides k or m: i.e., $k = 0$ or $m = 0$. Q.E.D.

20 Ring homomorphisms

(20.1) Definition *Let $R_i, +_i, *_i$, $i = 1, 2$, be two rings. A* (ring) homomorphism *from R_1 to R_2 is a map θ such that for every $a, b \in R_1$,*

$$\theta(a +_1 b) = \theta(a) +_2 \theta(b), \quad \textit{and} \quad \theta(a *_1 b) = \theta(a) *_2 \theta(b).$$

If θ is bijective it is an isomorphism *and the rings are* isomorphic, *written $R_1 \cong R_2$. An isomorphism from a ring to itself is called an* automorphism.

For example, the map $\theta\colon \mathbf{Z}[\sqrt{2}] \to \mathbf{Z}[\sqrt{2}]$, $a + b\sqrt{2} \mapsto a + b\sqrt{2}$, is an automorphism, since
(i) $(a+b\sqrt{2})+(c+d\sqrt{2}) = (a+c)+(b+d)\sqrt{2}) \mapsto (a+c)-(b+d)\sqrt{2}) = (a-b\sqrt{2})+(c-d\sqrt{2})$
(ii) $(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \mapsto (ac + 2bd) - (ad + bc)\sqrt{2} = (a - b\sqrt{2})(c - d\sqrt{2})$
(ii) $\theta \circ \theta$ is the identity map, so θ is bijective, an isomorphism.

(20.2) Definition *The* kernel *of a ring homomorphism $\theta\colon R_2 \to R_2$ is the set $\{x \in R_1 \colon \theta(x) = 0\}$.*

(20.3) Lemma *The image of a ring homomorphism $\theta\colon R_1 \to R_2$ is a subring of R_2.*

Proof. We need to show that the image is nonempty and closed under addition, negation, and multiplication. The first three facts need no proof since they apply to additive group homomorphisms and θ is one. For the fourth, we need to show that $\theta(x)\theta(y)$ is in the image whenever $x, y \in R_1$: but $\theta(x)\theta(y) = \theta(xy)$. Q.E.D.

(20.4) Lemma *If θ is a ring homomorphism then $\theta(x - y) = \theta(x) - \theta(y)$.*

Proof: $\theta(x - y) = \theta(x + (-y)) = \theta(x) + \theta(-y) = \theta(x) + (-\theta(y)) = \theta(x) - \theta(y)$. Q.E.D.

(20.5) Lemma *A ring homomorphism θ is injective if and only if its kernel is $\{0\}$.*

Proof. If: suppose the kernel is trivial and $\theta(x) = \theta(y)$. Then $\theta(x - y) = \theta(x) - \theta(y) = 0$, so $x - y = 0$ and $x = y$.
Only if: Suppose θ is injective. R.T.P. the kernel is trivial; so, given $\theta(x) = 0$, R.T.P. $x = 0$. But then $\theta(x) = \theta(0)$ so $x = 0$. Q.E.D.

21 Characteristic of a ring

(21.1) Definition *The* characteristic *of a ring R is the smallest positive integer c, if it exists, such that $cx = 0$ for all $x \in R$.*
If no such integer exists, the characteristic is zero.

For example, $\mathbf{Z}, \mathbf{Q}, \mathbf{H}, M(2, \mathbf{R})$ all have characteristic zero. $\mathbf{Z}_9 \times \mathbf{Z}_6$ has characteristic 18.

(21.2) Lemma *For any $k \in \mathbf{Z}$ and $x \in R$, $-(kx) = (-k)x$.*

Proof. $kx + (-k)x = (k + (-k))x = 0x = 0$.
Therefore $-(kx) + kx + (-k)x = -(kx)$, so $0 + (-kx) = -(kx)$, or $(-kx) = -(kx)$. Q.E.D.

(21.3) Lemma *Given x, y in a ring R and $n \in \mathbf{Z}$, $(nx)y = n(xy)$.*

Proof. The case $n \geq 0$ can be proved by induction on n and is left as an exercise. If $n < 0$, let $n = -k$. Assuming the result for k, $(kx)y = k(xy)$. Take the negative of each side:

$$-((kx)y) = -(k(xy)).$$

But $-((kx)y) = ((-(kx))y$ (Lemma 18.5 (b)). From the above lemma, $(-k)x = -(kx)$ and $(-k)(xy) = -(k(xy))$. Hence

$$((-k)x)y = (-(kx))y = -((kx)y) = -(k(xy)) = (-k)(xy).$$

That is, $(nx)y = n(xy)$. Q.E.D.

(21.4) Lemma *If R is a unital ring with unity e, then its characteristic is the least positive integer d, if it exists, such that $de = 0$. Otherwise the characteristic is zero.*

Proof. Let c be the characteristic of R.

If $d > 0$, then for all $x \in R$, $dx = d(ex) = (de)x$, by Lemma 21.3. But $(de)x = 0x = 0$, so $dx = 0$, and the characteristic c is nonzero and bounded by d. But c is not less than d, since if $c < d$ then $ce \neq 0$. Therefore $c = d$.

If $c > 0$ then $ce = 0$, so $d > 0$. Hence if $d = 0$ then $c = 0$. Q.E.D.

(21.5) Lemma *In a unital ring R with unity e, for any $m, n \in \mathbf{Z}$, (mn)e = (me)(ne).*

Proof: $(me)(ne) = m(e(ne))$ from the above Lemma. This equals $m(ne)$, which equals $(mn)e$ from the list of facts given above Lemma 18.5. Q.E.D.

(21.6) Lemma *If R is a unital ring of characteristic zero, then it has a subring isomorphic to $\mathbf{Z}$.*

Proof. Let $\theta\colon \mathbf{Z} \to R$ be the map taking n to ne.

Claim that θ is a homomorphism.

Since $(m + n)e = (me) + (ne)$, $\theta(m + n) = \theta(m) + \theta(n)$. Since $(mn)e = (me)(ne)$, (Lemma 21.5), $\theta(mn) = \theta(m)\theta(n)$.

Therefore θ is a homomorphism. Moreover, θ is injective, for the following reason.

its kernel is the set of those n such that $ne = 0$, which is $\{0\}$ since the characteristic is zero (Lemma 21.4). Therefore θ is injective (Lemma 20.5). The image Z of θ is a subring of R (Lemma 20.3). θ is a ring homomorphism which maps $\mathbf{Z}$ bijectively onto Z, and therefore Z is isomorphic to $\mathbf{Z}$. Q.E.D.

(21.7) Lemma *If R is a unital ring of characteristic $c > 0$, then it has a subring isomorphic to $\mathbf{Z}_c$.*

Proof. We consider the following map θ from $\mathbf{Z}_c$ to R: $n \mapsto ne$, for $0 \leq n < c$.

For clarity, we shall write $m + n$ for the sum as integers, and $(m + n) \mod c$ for the sum in $\mathbf{Z}_c$.

As in the previous lemma, given $m, n \in \mathbf{Z}$, $(m + n)e = (me) + (ne)$ and $(mn)e = (me)(ne)$. Given $m, n \in \mathbf{Z}_c$, let $k = (m+n) \mod c$ be their sum in $\mathbf{Z}_c$. Then $k = m+n+\ell c$ for some integer ℓ, so $ke = me+ne+(\ell c)e-me+ne+\ell(ce) == me+ne$, since $ce = 0$. That is, $\theta(k) = \theta(m)+\theta(n)$, so θ preserves addition in $\mathbf{Z}_c$.

Next let $k = (mn \mod c) = mn + \ell c$ for some ℓ. Then $ke = (mn + \ell c)e = (mn)e + (\ell c)e = (mn)e + \ell(ce) = (mn)e = (me)(ne)$, so θ is a ring homomorphism.

If $0 \leq n < c$ and $\theta(n) = 0$, then $ne = 0$ so n is a multiple of c, so $n = 0$, and θ is injective as in the previous lemma.

Thus θ maps $\mathbf{Z}_c$ bijectively onto its image in R, and its image is a subring of R isomorphic to $\mathbf{Z}_c$. Q.E.D.

22 Polynomials

(22.1) Definition *Let R be a commutative ring, x an 'indeterminate.' A* power series in x *is a formal (infinite) expression*

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots,$$

where a_r are called the coefficients *and x the* indeterminate. *One can write this power series as*

$$\sum_{r=0}^{\infty} a_r x^r \quad \text{or} \quad \sum_{r \geq 0} a_r x^r.$$

If all but finitely many coefficients are zero, it is called a polynomial, *and the largest n, if it exists, such that $a_n \neq 0$ is called its* degree. *If p is a polynomial then $\deg(p)$ denotes its degree.*

If all the coefficients are zero, then the degree is defined to be $-\infty$.

Alternatively, a polynomial can be defined as either

- The zero element of R, or
- $a_0 + a_1x + \ldots + a_nx^n$, where $a_n \neq 0$. In this case, n is its degree.

Addition and multiplication can be defined for any two power series, whether or not they are polynomials. The formulae are

$$\sum_{r=0}^{\infty} a_r x^r + \sum_{r=0}^{\infty} b_r x^r = \sum_{r=0}^{\infty} (a_r + b_r) x^r,$$

and

$$(\sum_{r=0}^{\infty} a_r x^r)(\sum_{r=0}^{\infty} b_r x^r) = \sum_{k=0}^{\infty} c_k x^k, \text{ where } c_k = \sum_{r+s=k} a_r b_s$$

Remark. In practice a polynomial is given as a finite sum with the zero coefficients omitted. The power series formulation avoids a good deal of mess.

Example. In $\mathbf{Z}[x]$, let $p = 2 + 3x$ and $q = 4 - 2x + 3x^2$. Then

$$p + q = 6 + x + 3x^2, \text{ and } pq = 8 + 8x + 9x^2.$$

(22.2) Lemma *If $p, q \in R[x]$, then* (i) *pq is also a polynomial, and* (ii) $\deg(pq) \leq \deg(p) + \deg(q)$. *If R is an integral domain, then* (iii) $\deg(pq) = \deg(p) + \deg(q)$.

Proof. (a) If p or q is zero, then so is pq, which is the zero polynomial, and $\deg(pq) = -\infty$ and $\deg(p) + \deg(q) = -\infty$. So otherwise assume $\deg(p) = m \geq 0$ and $\deg(q) = n \geq 0$. This proves (i) and (ii) when p or q is zero.

Otherwise, (b) write $p = \sum a_r x^r$ and $q = \sum b_s x^s$. For any $k \in \mathbf{N}$, the coefficient of x^k in pq is $\sum_{r+s=k} a_r b_s$. If $k > m + n$ and $r + s = k$, then either $r > m$ or $s > n$, so $a_r b_s = 0$. Therefore the coefficient of x^k in pq is zero if $k > m+n$, so pq is a polynomial of degree $\leq m+n$. This completes the proof of (i) and (ii).

If R is an integral domain, and p or q is zero, then the degree formula is an equation (part (a) above).

Otherwise, with m and n as in (b), consider $\sum_{r+s=m+n} a_r b_s$. If $r \neq m$, then either $r < m$ and $s > n$, or $r > m$. In either case, $a_r b_s = 0$. Hence the only possible nonzero term $a_r b_s$ with $r + s = m + n$ is $a_m b_n$. Since R is an integral domain, and $a_m \neq 0$ and $b_n \neq 0$, $a_m b_n \neq 0$, so $\deg(pq) = m + n = \deg(p) + \deg(q)$. Q.E.D.

(22.3) Lemma *$R[x]$ is an additive group.*

Proof. First, it is closed under addition. If $p = \sum a_r x^r$ and $q = \sum b_r x^r$ are polynomials, and $r > \max(\deg(p), \deg(q))$, then $a_r + b_r = 0$, so $\sum (a_r + b_r) x^r$ is a polynomial. The remainder of this proof applies equally well to power series as to polynomials.

Consider three power series $f = \sum_{r \geq 0} a_r x^r$, $g = \sum_{r \geq 0} b_r x^r$, and $h = \sum_{r \geq 0} c_r x^r$,

$$(f + g) + h = \sum (a_r + b_r) x^r + \sum c_r x^r =$$
$$\sum ((a_r + b_r) + c_r) x^r = \sum (a_r + (b_r + c_r)) x^r =$$
$$\sum a_r x^r + \sum (b_r + c_r) x^r$$
$$= f + (g + h).$$

Therefore addition is associative. Also,

$$f + g = \sum_r (a_r + b_r) x^r = \sum (b_r + a_r) x^r = g + f,$$

so addition is commutative. If g is the polynomial with all coefficients zero, then $f + g = \sum_r (a_r + 0) x^r = f$, so g functions as the additive identity. If we define $-f$ as $\sum_r (-a_r) x^r$ then $f + (-f) =$

$\sum(a_r - a_r)x^r = \sum 0x^r$, the additive identity, so the set of power series is an abelian group under addition.

Finally note that the zero element is a polynomial, and if power series p and q are polynomials, then so are $p + q$ and $-p$, so the polynomials form an abelian group under addition. Q.E.D.

(22.4) Lemma *$R[x]$ is a commutative semigroup under multiplication.*

Proof. By Lemma 22.2, $R[x]$ is closed under multiplication. The remainder of this proof can apply to power series as well as polynomials.

Given power series $f = \sum a_r x^r$, $g = \sum b_r x^r$, and $h = \sum c_r x^r$, let $fg = \sum d_k x^k$, $gh = \sum e_\ell x^\ell$, $(fg)h = \sum \alpha_m x^m$ and $f(gh) = \sum \beta_n x^n$.

$$\alpha_m = \sum_{k+t=m} d_k c_t = \sum_{k+t=m} (\sum_{r+s=k} a_r b_s)c_t =$$
$$\sum_{k=0}^{m} \sum_{r+s=k} a_r b_s c_{m-k} = \sum r + s + t = m a_r b_s c_t.$$

$$\beta_n = \sum_{r+\ell=n} a_r e_\ell =$$
$$\sum_{\ell=0}^{n} a_{n-\ell} \sum_{s+t=\ell} b_s c_t = \sum_{r+s+t=n} a_r b_s c_t.$$

Thus, $\alpha_m = \beta_m$ for each m, so $(fg)h = f(gh)$, and multiplication is associative.

Multiplication is commutative: $fg = \sum_k (\sum_{r+s=k} a_r b_s)x^k$, $gf = \sum_k (\sum_{s+r=k} b_s a_r)x^k$. These are identical, so multiplication is commutative. Q.E.D.

(22.5) Lemma *$R[x]$ is a commutative ring.*

Proof. We know that it is an abelian group under addition and a commutative semigroup under multiplcation. We need to check the distributive laws, and because it is commutative, we need only check one of them. This applies to power series as well as polynomials.

So let $f = \sum a_r x^r$, $g = \sum b_s x^s$, and $h = \sum c_t x^t$.

$$f(g + h) = \sum_k (\sum_{r+s=k} a_r(b_s + c_s))x^k = \sum_k (\sum_{r+s=k} a_r b_s)x^k + \sum_k (\sum_{r+s=k} a_r c_s)x^k = (fg) + (fh)$$

Q.E.D.

(22.6) Lemma *R is isomorphic to a subring of $R[x]$.*

Proof. The map $R \to R[x]$; $a \mapsto ax^0$, is a homomorphism, since $a + b \mapsto ax^0 + bx^0$ and $ab \mapsto abx^0$. If ax^0 is the zero polynomial then $a = 0$, so this map is injective and it takes R isomorphimcally onto the subring $\{ax^0 : a \in R\}$. Q.E.D.

(22.7) Lemma *$R[x]$ is unital if R is. (Easy. The unit is ex^0.)* ∎

(22.8) Lemma *If R is an integral domain, then so is $R[x]$.*

Proof. Suppose that $p, q \neq 0$, so $\deg(p) \geq 0$ and $\deg(q) \geq 0$, $\deg(pq) = \deg(p) + \deg(q) \geq 0$. Q.E.D.

23 Division algorithm for polynomials over a field

Recall how to divide polynomials.

	x^2	$+$	$4x$	$+$	2				
$x^2 - x + 2$	x^4	$+$	$3x^3$			$+$	x	$-$	5
	x^4	$-$	x^3	$+$	$2x^2$				
			$4x^3$	$-$	$2x^2$	$+$	x		
			$4x^3$	$-$	$4x^2$	$+$	$8x$		
					$2x^2$	$-$	$7x$	$-$	5
					$2x^2$	$-$	$2x$	$+$	4
							$-5x$	$-$	9

Thus $x^4 + 3x^3 + x - 5 = (x^2 - x + 2)(x^2 + 4x + 2) + (-5x - 9)$. More generally,

(23.1) Theorem *If $p, d \in F[x]$ are polynomials over a field F and $d \neq 0$, then there exist unique polynomials d, r such that $p = qd + r$ and $\deg(r) < \deg(d)$.*

Proof: existence by induction on $\deg(p)$, assuming d is fixed. If $\deg(p) < \deg(d)$ we can take $r = p$ and $q = 0$. This covers the base case.

For the inductive step, suppose that $p(x) = a_n x^n + \ldots$, and $d = b_m x^m + \ldots$, where $n = \deg(p)$ and $m = \deg(d)$; $n \geq m$. Consider

$$p - \frac{a_n}{b_m} x^{n-m} d.$$

The degree of p is n, as is that of $x^{n-m}d$. Hence the above polynomial has degree at most n. However, the coefficient of x^n is $a_n - a_n = 0$, so the degree of the result is $< \deg(p)$.

By the inductive hypothesis, there exist polynomials q', r where $\deg(r) < \deg(d)$ and

$$p - \frac{a_n}{b_m} x^{n-m} d = q'd + r,$$

so

$$p = (\frac{a_n}{b_m} x^{n-m} + q')d + r,$$

i.e., $p = qd + r$ as required.

Uniqueness. Suppose $qd + r = q'd + r'$ where $\deg(r) < \deg(d)$ and $\deg(r') < \deg(d)$. Then $(q - q')d = r - r'$. If $q \neq q'$ then the left-hand side has degree $\geq \deg(d)$. But the right-hand side has degree $< \deg(d)$, which is false. Hence $q = q'$ and it follows that $r = r'$. Q.E.D.

24 Factorising polynomials

This will mostly be about factorising polynomials over **Q**.

(24.1) Definition *A polynomial is* monic *if the coefficient of its highest-degree term is 1 (the unity in F).*

Needless to say, if p, q are polynomials, then $p|q$ means there exists another polynomial d such that $q = dp$.

(24.2) Theorem *Let F be a field and $a, b \in F[x]$, not both zero. Then there exists a unique* monic *polynomial $d(x)$ such that* (a) *$d|a$ and $d|b$, and* (b) *If $c \in F[x]$ and $c|a$ and $c|b$ then $c|d$.*

Moreover, **(c)** *there exist polynomials r and s such that $d = ra + sb$.*

Proof: existence. Let d be a monic polynomial of minimum degree representable as $ra + sb$ for polynomials r and s.

(Since a and b are not both zero and F is a field there exist monic polynomials of the form $ra+sb$, so d exists by the well-ordering priciple).

Divide a by d: $a = qd + g$, say. If $g \neq 0$ then $g = a - qd$ would be of the form $ta + ub$, nonzero, of lower degree than d, and by scaling it can be made monic, contadicting the definition of d. Hence $d|a$ and similarly $d|b$.

If $c|a$ and $c|b$ then $c|ra + sb = d$. This concludes the proof of existence.

Uniqueness: Suppose d and d' both satisfy (a) and (b). Then $d|d'$, so $d' = ud$ for some polynomial u. Comparing highest-degree terms, u must be monic. Also $d'|d$, so $d = vd'$ for some monic polynomial v. Then $d = uvd/$ But $d \neq 0$ and $F[x]$ is an integral domain, so $(1 - uv)d = 0$, so $uv = 1$. Hence u and v are constants, and monic, so $u = v = 1$ and $d = d'$. Q.E.D.

(24.3) Definition *When a, b are polynomials over a field F, not both zero, and d is the monic polynomial introduced above, we write $d = \gcd(a, b)$.*

Remark. The gcd and the polynomials r and s such that $\gcd(a, b) = ra + sb$ can be calculated by a variant of Euclid's gcd algorithm.

Example: $\gcd(x^5 - x^4 + 2x^3 + x + 1, x^3 - x^2 + x - 1)$.

	x^2	$+$	1										
$x^3 - x^2 + x - 1$	x^5	$-$	x^4	$+$	$2x^3$				$+$	x	$+$	1
	x^5	$-$	x^4	$+$	x^3	$+$	x^2					
					x^3	$+$	x^2	$+$	x	$+$	1
					x^3	$-$	x^2	$+$	x	$-$	1
							$2x^2$			$+$	2

$q_1 = x^2 + 1, r_1 = 2(x^2 + 1), a = q_1 b + r_1$.

	$x/2$	$-$	$1/2$				
$2x^2 + 2$	x^3	$-$	x^2	$+$	x	$-$	1
	x^3			$+$	x		
			$-x^2$			$-$	1
			$-x^2$			$-$	1
							0

Assuming that $\deg(a) \geq \deg(b)$, one starts with $p_1 = a$ and $p_2 = b$, and repeatedly divide p_{j+1} into p_i with quotient q_i and remainder r_i, then set $p_{i+1} = r_i$. Continue so long as $p_{i+1} \neq 0$. The last nonzero p_i, after scaling it to make it monic, is the gcd.

In this example, $p_1 = x^5 - x^4 + 2x^3 + x + 1$, $p_2 = x^3 - x^2 + x - 1$, $p_3 = 2(x^2 + 2)$, $p_4 = 0$. Scale p_3 to get the gcd, $x^2 + 1$.

$p_3 = p_1 - (x^2 + 1)p_2$, and p_3 (rescaled) is the gcd. The rescaling is dividng by 2. So $x + 2 + 1 = a/2 - (x^2 + 1)/2b$.

This gives us polynomials $r = 1/2, s = -(x^2 + 1)/2$, so one can express th gcd as $ra + sb$.

(24.4) Lemma (evaluation map). *Given an element c of a commutative ring r, there exists a unique homomorphism $\theta\colon R[x] \to R$ which fixes R, that is, $\theta(a) = ax^0$ for each $a \in R$, and takes x to c.*

Proof: uniqueness. Since θ is a homomorphism, for any polynomial $p = a_0 + a_1x + \ldots a_mx^m$,

$$\theta(p) = \theta(a_0) + \theta(a_1)\theta(x) + \ldots \theta(a_m)(\theta(x))^m \quad = \quad a_0 + a_1c + \ldots a_mc^m. \tag{24.5}$$

This fixes θ uniquely.

Existence. Again, if θ is the map defined as in the above equation 24.5, it clearly fixes R and maps x to c. we need show that θ is a homomorphism

Let a polynomial p be written as a power series $\sum a_rx^r$, where $a_r = 0$ for $r > m$. Let q be another polynomial: write it as $\sum b_sx^s$, again as a power series.

The map θ is additive: $\theta(p) + \theta(q) = \sum a_rc^r + \sum b_sc^s = \sum(a_r + b_r)c^r = \theta(p + 1)$.

The map θ is multiplicative:

$$\theta(p)\theta(q) = (\sum a_rc^r)(\sum b_sc^s) =$$
$$\sum_{r,s} a_rb_sc^{r+s} = \sum_k (\sum_{r+s=k} a_rb_s)c^k = \theta(pq).$$

Therefore θ is a homomorphism. Q.E.D.

(24.6) Notation *Rather than $\theta(f)$ we wriet $f(c)$.*

(24.7) Corollary *Let F be a field. Given $c \in F$ and $f \in F[x]$, the remainder on dividing f by $x - c$ is $f(c)$.*

Proof. We know there exist unique polynomials q and r where $f = (x - c)q + r$ and $\deg(r) < \deg(x - c)$. Apply the above evaluation map θ. Since this is a homomorphism,

$$\theta(f) = (c - c)\theta(q) + \theta(r)$$

But r has degree ≤ 0, so $\theta(r) = r$. Therefore $r = f(c)$. Q.E.D.

(24.8) Corollary *Given $c \in F$ and a polynomial f over a field F, $(x - c)$ divides f if and only if $f(c) = 0$. (Trivial.)* ∎

(24.9) Definition *Two polynomial s f, g over a field F are* associates *of $f = cg$ for some $c \neq 0$ in F.*

Any polynomial of degree $geq1$ has many divisors: all nonzero field elements, all its associates.

(24.10) Definition *A polynomail $f(x) \in F[x]$ (F a field) is* irreducible *if $\deg(f) \geq 1$ and if $f = gh$, then g is a constant and h an associate of f or vice-versa.*

(24.11) Lemma *If $p \in F[x]$ (F a field) is irreducible and $p|ab$, $a, b \in F[x]$, then $p|a$ or $p|b$.*

Proof. *Suppose $p \not| a$: R.T.P. $p|b$. Since p is irreducible and $\gcd(p, a)$ divides p and is not an associate of p (since $p \not| a$), $\gcd(p, a) = 1$ (the unity of F). There exist polynomials r, s such that $rp + sa = 1$.*

Then $rpb + sab = b$, and p divides the left-hand side, so $p|b$. Q.E.D.

From this, as with integer factorisation, te following can be proved, though we omit a proof.

(24.12) Proposition (polynomial unique factorisation theorem). *Let F be a field. Every polynomal f of degree ≥ 1 in $F[x]$ can be expressed essentially uniquely in the form*

$$f = ap_1p_2 \cdots p_k$$

where $a \in F$ and p_j are irreducible monic polynomials. The factorisation is unique except for the order of factors. ∎

25 Gauss's Lemma and Eisenstein's Criterion

This section is about factorising polynomials over $\mathbf{Z}[x]$. We first prove that a polynomial in $\mathbf{Z}[x]$ is irreducible in $\mathbf{Z}[x]$ if and only if it is irreducible in $\mathbf{Q}[x]$. The 'only if' part is the interesting part.

(25.1) Definition *A polynomial $a_0 + a_1x + \ldots + a_nx^n$ in $\mathbf{Z}[x]$ is* primitive *if it is nonzero and $\gcd(a_0, a_1, \ldots, a_n) = 1$.*

(25.2) Lemma *Given a nonzero polynomial $f = a_0 + a_1x + \ldots + a_nx^n$ in $\mathbf{Z}[x]$, let*

$$d = \gcd(a_0, a_1, \ldots, a_n).$$

Let $f' = f/d$. Then f' is a primitive polynomial in $\mathbf{Z}[x]$.

Proof. Write a'_r for a_r/d — so $a_r = da'_r$, since $d|a_r$ for all r. Then $f' = \sum a'_r x^r$ and $f = dp'$.

To show that f' is primitive, suppose that c is a positive integer dividing each a'_r. R.T.P. $d = 1$.For each r, dc divides a'_r, so dc divides each a_r, so dc divides d and $c = 1$. Q.E.D.

(25.3) Lemma (Gauss's Lemma). *The product of two primitive polynomials (in $\mathbf{Z}[x]$) is primitive.*

Proof. Let $f = \sum a_r x^r$ and $g = \sum b_s x^s$ be polynomials over $\mathbf{Z}$, and suppose they are primitve. Write $fg = \sum c_k x^k$ where $c_k = \sum) r + s = k a_r b_s$. Let p be any prime: R.T.P. there exists a k such that $p \not| c_k$.

Let i be the *smallest* index such that a_i is not divisible by p: i exists because f is primitive. Let j be the smallest index, which exists because g is primitive, such that b_j is not dividible by p. Let $k = i + j$. Consider $c_k = \sum_{r+s=k} a_r b_s$.

If $r < i$ then $p | a_r b_s$. If $r > i$ then $s < j$ and $p | a_r b_s$. Hence p divides every term in the sum except possibly $a_i b_j$. But p does not divide $a_i b_j$ since it divides neither a_i nor b_j. Hence p does not divide c_k. Q.E.D.

(25.4) Lemma *Suppose that $f \in \mathbf{Z}[x]$ is primitive and $a, b \in \mathbf{Z}$. If $b | af$ then $b | a$.*

Proof. Suppose $f = a_0 + a_1 x + \ldots + a_n x_n$. The GCD of the coefficients can be expressed in the form

$$r_0 a_0 + r_1 a_1 + \ldots + r_n a_n.$$

Hence this expression is 1. If $b | a a_j$ for each j, then

$$b | a r_0 a_0 + a r_1 a_1 + \ldots + a r_n a_n = a,$$

so $b | a$. Q.E.D.

(25.5) Corollary *A polynomial (of degree ≥ 1) in $\mathbf{Z}[x]$ is irreducible over $\mathbf{Z}[x]$ if and only if it is irreducible over $\mathbf{Z}[x]$.*

Proof. Let a be the GCD of the coefficents in the polynomial; then the polynomial can be written as af where f is primitive.

Suppose af is reducible over $\mathbf{Q}$: $af = g_1 h_1$ where g_1, h_1 are polynomials of positive degree in $IQ[x]$.

Let c' be the LCM of the coefficients in g_1, so $g_1 = g_2/c'$ where $g_2 \in \mathbf{Z}[x]$. Let b' be the GCD of the coefficients in g_2, so $g_2 = b'g$ where g is primitive. Let b/c be the reduced form of the fraction b'/c' so $\gcd(b, c) = 1$. Thus $g_1 = bg/c$. Similarly, $h_1 = dh/e$ where h is primitive. Then

$$acef = bdgh.$$

From the above lemma, $bd | ace$, and by Gauss's Lemma, gh is primitive, so by the above lemma, $ace | bd$. Therefore $ace = \pm bd$. If $ace = -bd$, then g can be replaced by $-g$ and b by $-b$, so $ace = bd$.

Since c divides bd, c divides d, i.e., d/c is an integer. Similarly b/e is an integer, and $a = (b/e)(d/c)$. Thus,

$$af = (b/e)g(d/c)h$$

expresses af as a product of two polynomials in $\mathbf{Z}[x]$. Q.E.D.

(25.6) Theorem (Eisenstein's Criterion). *Let $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ be a polynomial in $\mathbf{Z}[x]$, of positive degree. Suppose that there exists a prime p such that*

$$p | a_r, r = 0, \ldots, n-1; \quad p \not| a_n; \quad p^2 \not| a_0.$$

Then f is irreducible over $\mathbf{Z}[x]$, hence also over $\mathbf{Q}[x]$.

Proof. Suppose

$$a_0 + \ldots + a_n x^n = (b_0 + b_1 x + \ldots + b_k x^k)(c_0 + c_1 x + \ldots + c_\ell x^\ell)$$

where b_j and c_j are integers. Since p divides a_0 but p^2 does not, and $a_0 = b_0 c_0$, p divides exactly one of b_0 and c_0. Without loss of generality, p does not divide b_0 and divides c_0.

Prove by induction on j that p divides c_j, for $0 \leq j \leq \ell$. Assuming the result for j with $j < \ell$, consider $a_{j+1} = \sum_{r+s=j+1} b_r c_s$.

This is $b_0 c_{j+1} + \sum_{r=1}^{j+1} b_r c_{j+1-r}$. Every term in the latter sum is divisible by p, and so is a_{j+1} since $j + 1 \leq \ell < \ell + k = n$. Therefore $p | b_0 c_{j+1}$ and $p \not| b_0$ so $p | c_{j+1}$. This concludes the inductive proof.

Therefore p divides all coefficients c_j, so $p | b_k c_\ell$, i.e., $p | a_n$. But this is false, so the factorisation does not exist. Q.E.D.

26 Ring homomorphisms and ideals

(26.1) Definition *A* left ideal *in a ring R is a subring I such that for every $a \in I$ and $r \in R$, $ra \in I$.*
A right ideal *in a ring R is a subring I such that for every $a \in I$ and $r \in R$, $ar \in I$.*
A (2-sided) ideal *I in R is a subring which is both a left and a right ideal.*

Example. In $M(2, \mathbf{R})$,

$$\left\{ \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} : x, y \in \mathbf{R} \right\}$$

is a left ideal but not a right ideal.

Example. Every additive subgroup of $\mathbf{Z}$ is an ideal, because every such subgroup is of the form $\langle g \rangle$, the set of all multiples of g, and if a is a multiple of g then so is $ra = ar$ for every integer r.

For any commutative nunital ring R and any $a \in R$, the set $ra \colon r \in R$ is an ideal containing a.

A field F has no nontrivial ideals, that is, if $I \neq \{0\}$ is an ideal in F, I contains some nonzero x, so it contains $x^{-1}x$, hence contains $rx^{-1}x = r$ for every $r \in F$, so $I = F$.

(26.2) Lemma *Let $\theta \colon R_1 \to R_2$ be a ring homomorphism. Then its kernel is an ideal in R_1.* ∎

(26.3) Definition *Given a ring R and an ideal I in R, there are well-defined addition and multiplication operations on the cosets $r + I$ of I:*

$$(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I;$$
$$(r_1 + I)(r_2 + I) = (r_1 r_2) + I.$$

With these definitions, the set R/I of cosets of I form a ring.

Proof. Since $R, +$ is an abelian group, addition as given here is well-defined and makes R/I an abelian group under addition.

To show it is well-defined under multiplication, suppose $r_1 + I = r_1' + I$ and $r_2 + I = r_2' + I$. Then $r_1 - r_1' \in I$ $r_2 - r_2' \in I$. R.T.P. $r_1 r_2 + I = r_1' r_2' + I$.

Since I is a (2-sided) ideal, $(r_1 - r_1')r_2 + r_1'(r_2 - r_2') \in I$. That is, $r_1r_2 - r_1'r_2' \in I$, so $r_1r_2 + I = r_1'r_2' + I$, as required.

$$(r_1 + I)((r_2 + I)(r_3 + I)) = (r_1 + I)(r_2r_3 + I) = r_1(r_2r_3) + I$$
$$= (r_1r_2)r_3 + I = (r_1r_2 + I)(r_3 + I) = ((r_1 + I)(r_2 + I))(r_3 + I),$$

so multiplication is associative.

$$(r_1 + I)((r_2 + I) + (r_3 + I)) = (r_1 + I)((r + 2 + r_3) + I) = (r_1(r_2 + r_3) + I) =$$
$$((r_1r_2) + (r_1r_3)) + I = (r_1r_2 + I) + (r_1r_3 + I) =$$
$$(r_1 + I)(r_2 + I) + (r_1 + I)(r_3 + I),$$

so one of the distributive laws holds. The other distibutive law

$$((r_1 + I) + (r_2 + I))(r_3 + I) = (r_1 + I)(r_3 + I) + (r_2 + I)(r_3 + I)$$

is proved in the same way. Q.E.D.

(26.4) Definition *R/I is called a quotient ring, the quotient of R by I.*

(26.5) Theorem (isomorphism theorem for rings). *Let $\theta\colon R_1 \to R_2$ be a ring homomorphism, and let $I = \ker(\theta)$.*

Then there is a well-defined map $\overline{\theta}\colon R_1/I \to R_2$ which carries R_1/I isomorphically onto the image of θ.

Proof. Define $\overline{\theta}(a + I) = \theta(a)$.

To show $\overline{\theta}$ is well-defined, suppose that $a + I = b + I$: R.T.P. $\theta(a) = \theta(b)$.

Since $a - b \in I$, $\theta(a - b) = 0$, so $\theta(a) - \theta(b) = 0$, so $\theta(a) = \theta(b)$, as required.

Given $a, b \in R_1$,

$$\overline{\theta}((a + I) + (b + I)) = \overline{\theta}((a + b) + I) =$$
$$\theta(a + b) = \theta(a) + \theta(b) = \overline{\theta}(a + I) + \overline{\theta}(b + I),$$

so $\overline{\theta}$ is additive.

$$\overline{\theta}((a + I)(b + I)) = \overline{\theta}(ab + I) =$$
$$\theta(ab) = \theta(a)\theta(b) = \overline{\theta}(a + I)\overline{\theta}(b + I),$$

so $\overline{\theta}$ is multiplicative, a homomorphism.

For any y in the image of θ, $y = \theta(x)$ for some $x \in R_1$, so $y = \overline{\theta}(x + I)$, that is, y is in the image of $\overline{\theta}$, Conversely, if y is in the image of $\overline{\theta}$ then it is in the image of θ, so $\overline{\theta}$ takes R_1/I surjectively onto the image of θ.

Finally, if $\overline{\theta}(a + I) = \overline{\theta}(b + I)$, then $\theta(a) = \theta(b)$, so $\theta(a - b) = \theta(a) - \theta(b) = 0$, so $a - b \in I$, so $a + I = b + I$. Therefore $\overline{\theta}$ is injective, so it carries R_1/I isomorphically onto the image of θ. Q.E.D.

Example. For any positive integer b, the map $\mathbf{Z} \to \mathbf{Z}_b$ taking m to $m \mod n$ is a *ring* homomorphism, and $\mathbf{Z}_n \cong \mathbf{Z}/(n)$, where (n) denotes the ring ideal generated by n — identical as a set with the additive subgroup $\langle n \rangle$ of $\mathbf{Z}$.

27 Principal ideal domains

(27.1) Definition *Let R be a ring, $a \in R$, The smallest ideal containing a is called the ideal generated by a and written (a).*

If R is commutative and unital, $(a) = \{ra\colon r \in R\}$.

(27.2) Definition *An integral domain D is called a* principal ideal domain *if for every ideal I in D, $I = (a)$ for some $a \in I$.*

Examples. Z is a principal ideal domain.
If F is a field, then $F[x]$ is a principal ideal domain.
$\mathbf{Z}[x]$ is *not* a principal ideal domain.

(27.3) Lemma *If F is a field then $F[x]$ is a principal ideal domain.*

Proof. Let I be an ideal in $F[x]$. If $I = \{0\}$ then $I = (0)$. Otherwise, let d be a nonzero polynomial of minimal degree in I.

For any $f \in I$, $f = qd + r$ where $\deg(r) < \deg(d)$, by the division algorithm.

Since $r = f - qd \in I$ and $\deg(r) < \deg(d)$, r must be zero, so $f = qd$, so $f \in (d)$. Therefore $I = (d)$. Q.E.D.

28 Quotient rings of $F[x]$

Throughout this section F will be a field.

(28.1) Theorem *Suppose that p is an irreducible polynomial in $F[x]$. Then $F[x]/(p)$ is a field containing a subfield isomorphic to F, over which the polynomial p has a linear factor.*

Proof. Since p is irreducible, p has positive degree. Without loss of generality p is monic.

Let 1 be the unity of F. For any polynomial f, $(1 + (p))(f + (p)) = f + (p)$, and $F[x]/(p)$ is commutative, so $F[x]/(p)$ is unital with unity $1 + (p)$.

Let $f + (p)$ be an element of the quotient ring. If $\gcd(p, f) \neq 1$ then $\gcd(p, f) = p$, p divides f, and $f \in (p)$. Otherwise $\gcd(p, f) = 1$, so there exist polynomials r and s such that $rf + sp = 1$, so $rf \in 1 + (p)$ Then $r + (p)$ is the reciprocal of $f + (p)$ in $F[x]/(p)$. Therefore $F[x]/(p)$ is a field.

Since p has positive degree, the only constant element of (p) is 0. Therefore, for any $a, b \in F$, if $a \neq b$ then $a - b \notin (p)$, so $a - b + (p) \neq (p)$, so $a + (p) \neq b + (p)$. Therefore the quotient map acts injectively on constant polynomials, i.e., elements of F, so F is mapped isomorphically to a subring of $F[x]/(p)$, which is a subfield.

Let K denote $F[x]/(p)$. Since $x + I \in K$, to write $K[x]$ would be ambiguous, so we choose a new indeterminate y and consider the ring $K[y]$ of polynomials with indeterminate y and coefficients in K.

If $p(x) = a_0 + a_1x + \ldots + a_mx^m$ in $F[x]$, there is a similar polynomial $p(y) = a_0 + a_1y + \ldots + a_my^m$ in $K[y]$,

To show that $p(y)$ has a linear factor in $K[y]$, it is enough to show, by the factor theorem (Corollary 24.8), that $p(y)$ has a zero in K: we shall prove that $x + (p)$ is a zero for $p(y)$ in K.

One can easily prove by induction that for any polynomial $f(y) \in K[y]$, $f(x+(p)) = f(x)+(p)$.

In particular, $p(x + (p)) = p(x) + (p) = (p)$, the zero in the quotient ring $F[x]/(p)$. In other words, $x + (p)$ is a root of the polynomial $p(y)$ in $K[y]$, as claimed. Q.E.D.

(28.2) Lemma *If $f \in F[x]$ is reducible, then $F[x]/(f)$ is not an integral domain.*

Proof. Write $f = gh$ where g, h are polynomials of positive degree. Since all nonzero elements of (f) have degree at least that of f, $g + (f)$ and $h + (f)$ are both nonzero (i.e., different from (f)).

Then $(g + (f))(h + (f)) = gh + (f) = (f)$, so $g + (f)$ and $h + (f)$ are zero divisors. Q.E.D.

29 Dimension of extension fields

For simplicity, most fields considered will be subfields of $\mathbf{C}$. This is justified by the is the so-called

(29.1) Proposition (Fundamental theorem of algebra.) *Every non-constant polynomal in $\mathbf{C}[x]$ splits into a product of linear factors in $\mathbf{C}[x]$. (Proof omitted.)*

We require some linear algebra.

(29.2) Let K be a field, In the context of vector spaces, elements of K are called *scalars*, and elements of the vector space are called *vectors*. A *vector space* V over K is an abelian group $V, +$ together with a law of *multiplicication by scalars*: if $\alpha \in K$ and $xinV$ then αx is another vector in V. There are other axioms, such as $1x = x$, $\alpha(u + v) = (\alpha u) + (\alpha v)$, etcetera. Details are omitted.

Examples.

- The 'cartesian plane' $\mathbf{R}\times\mathbf{R}$, with $(x_1, y_1)+(x_2, y_2) = (x_1+x_2, y_1+y_2)$ and $\alpha(x, y) = (\alpha x, \alpha y)$.
- $\mathbf{R}^n$ over $\mathbf{R}$.
- K^n over K, where K is a field.
- $\mathbf{R}[x]$ with constant polynomials treated as scalars.
- Polynomials of degree ≤ 2 in $\mathbf{R}[x]$.
- $\mathbf{Q}[\sqrt{2}]$, viewed as a vector space over $\mathbf{Q}$.
- $\mathbf{R}$, viewed as a vector space over $\mathbf{Q}$.
- If K is a subfield of L then we call L an extension field over K. It can be considered as a vector space over K.

(29.3) Definition *A* linear isomorphism $\theta\colon V \to W$ *of vector spaces over K is a bijective map which preserves addition and multiplication by scalars.*

Equivalently, for all $v_1, v_2 \in V$ and $\alpha_1, \alpha_2 \in K$,

$$\theta(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1\theta(v_1) + \alpha_2\theta(v_2).$$

If $W = K^n$ (K being the field), θ is called a coordinate system.

Coordinate systems are usually related to what are called *bases*:

(29.4) Definition *Let V be a vector space over a field K. An ordered list $x_1, \ldots, x_n \in V$ of vectors is called a* finite basis *if for every $v \in V$ there exists a* unique *list of scalars $(\alpha_1, \ldots \alpha_n)$ such that*

$$v = \alpha_1 x_1 + \ldots \alpha_n x_n$$

(29.5) Lemma *If $x_1, \ldots, x_n$ is a finite basis, then with v and $\alpha_1, \ldots, \alpha_n$ as in tbe above definition, the map*

$$\theta \colon v \mapsto (\alpha_1, \ldots, \alpha_n)$$

is a coordinate system.

Conversely, if $\theta \colon V \to K^n$ is a coordinate system, $x_1 = \theta^{-1}((1, 0, \ldots, 0)), x_2 = \theta^{-1}((0, 1, 0, \ldots, 0))$, etcetera, then $x_1, \ldots, x_n$ is a basis from which θ can be defined in this way.

For example, $(1, 0), (0, 1)$ is a basis for $\mathbf{R}^2$. So is $(1, 2), (3, 4)$.
Question: Is $(1, 2, 3), (4, 5, 6), (7, 8, 9)$ a basis for $\mathbf{R}^3$?

(29.6) Definition *A vector space V is* finite-dimensional *if there exists a finite basis for V.*

Infinite-dimensional vector spaces exist, but are not of interest here.

(29.7) Lemma *If X, Y are (finite) bases for V and $X \subseteq Y$ then $X = Y$. (Proof: exercise.)*

(29.8) Lemma (Exchange Lemma.) *If X and Y are both (finite) bases for a vector space V. and $x \in X$, then there exists a vector y in Y such that $Y \backslash \{y\} \cup \{x\}$ is also a basis for V. (Proof not very difficult, but omitted.)*

(29.9) Corollary *If V is a finite-dimensional vector space, then any two bases contain exactly the same number of vectors.*

Proof. Notice that $Y \backslash \{y\} \cup \{x\}$ has the same cardinality as Y. Repeating the exhange process sufficiently often, we end up with a basis Y' which has the same cardinality as Y and contains X. From Lemma 29.7, $|X| = |Y'|$. Since $|Y| = |Y'|$, $|Y| = |X|$. Q.E.D.

(29.10) Definition *The* dimension *of a finite-dimensional vector space is the number of elements in any (finite) basis.*

If $K \subseteq L$ are fields, and L is considered as a vector space over K, then the dimension of L is written $[L : K]$ and called the degree *of L over K.*

Remark. $[\mathbf{R} : \mathbf{Q}] = \infty$.

(29.11) Definition *Let K be a subfield of $\mathbf{C}$, say.*

If u is an element of $\mathbf{C}$, then $K[u]$ denotes the subring of $\mathbf{C}$ consisting of expressions $f(u)$, where $f(x) \in K[x]$.

If $p(u) = 0$ for some polynomial $p \in K[x]$. then u is called algebraic over K

(29.12) Lemma *For any subfield K of $\mathbf{C}$ and any $u \in \mathbf{C}$,*
(i)*the set $\{p \in K[x] \colon p(u) = 0\}$ is an ideal in $K[x]$.*
(ii) *This ideal is nontrivial if and only if u is algebraic over K.*
(iii) *If u is algebraic over K then there exists a unique monic irreducible polynomial p such that $p(u) = 0$*
(iv) *If u is algebraic over K then $K[u]$ is a finite-degree extension field of K, and $[K[u] : K]$ is the degree of this unique polynomial p.*

Proof. (i) This set contains the zero polynomial, and for any polynomials $f(x), g(x), h(x) \in K[x]$, $f(u) = 0$ and $g(u) = 0$ implies $(f + g)(u) = 0$ and $(hf)(u) = 0$, so the set is an ideal in $K[x]$.

(ii) Clearly, this ideal, call it I, contains no constant polynomial except 0. Thus I is nontrivial if and only if it contains some nonconstant polynomial, or equivalently u is algebraic over K.

(iii) Suppose that u is algebraic over K, let I be the ideal as discussed, and let p be the unique monic polynomial generating I. If $p = qr$ where q and r are monic polynomials of lower degree, then $q * u) = 0$ or $r(u) = 0$ so $q \in I$ or $r \in I$, which is impossible. Therefore p must be irreducible.

If q is any irreducible monic polynomial such that $q(u) = 0$, then p divides the irreducible polynomial q, so $p = q$, so there is exactly one such irreducible monic polynomial.

(iv) Write the polynomial p as

$$p(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0.$$

If $n = 1$ then $u = a_0$ belongs to K and $K[u] = K$. In general, since $p(u) = 0$, u^n can be expressed as a polynomial of degree at most $n - 1$ in $K[u]$, and hence, by a simple inductive argument, for any $k \geq n$ u^k can be expressed as a polynomial of degree at most $n - 1$ in $K[u]$. Thus, every element v of $K[u]$ can be expressed in the form

$$v = \alpha_0 + \alpha_1 u + \alpha_2 u^2 + \ldots + \alpha_{n-1}u^{n-1}.$$

Furthermore, if v could be expressed differently as

$$v = \beta_0 + \beta_1 u + \beta_2 u^2 + \ldots + \beta_{n-1}u^{n-1},$$

then the polynomial

$$q(x) = (\alpha_{n-1} - \beta_{n-1})x^{n-1} + (\alpha_{n-2} - \beta_{n-2})x^{n-2} + \ldots (\alpha_0 - \beta_0)$$

would have the property that $q(u) = v - v = 0$, so q would belong to I, and $\deg(q) < n$, but $q \neq 0$, which is impossible. In other words,

$$1, u, u^2, \ldots, u^{n-1}$$

is a basis for $K[u]$ as a vector space over K, so $[K[u] : K] = n$.

Finally, to show that $K[u]$ is a field, we need show that any nonzero element has a reciprocal in $K[u]$. Let v be any nonzero element of $K[u]$. There exists a polynomial $f(x)$ of degree at most $n - 1$ such that $v = f(u)$. Since $\deg(f) < n$ and f is nonzero, $p(x)$ does not divide $f(x)$. Therefore the $gcd(p, f) = 1$. Therefore there exist polynomials r and s such that $r(x)p(x) + s(x)f(x) = 1$. Then $r(u)p(u) + s(u)f(u) = 1$, i.e., $s(u)v = 1$, so $s(u)$ is the reciprocal of v. Q.E.D.

(29.13) Definition *The polynomial introduced above is called the* minimum polynomial *for u over K.*

(29.14) Definition *Let V be a vector space over a field K. A list $x_1, \ldots, x_k$ of vectors in V is* linearly independent *if the only possible list of scalars $\alpha_1, \ldots, \alpha_k$ such that*

$$\alpha_1 x_1 + \ldots + \alpha_k x_k = 0$$

is $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

(29.15) Lemma (Another exchange lemma.) (i) *Suppose that X is a linearly independent list of vectors and Y a finite basis in a vector space V. Then for every $x \in X$ there exists a $y \in Y$ such that $Y \backslash \{y\} \cup \{x\}$ is another basis.*
(ii) *X can be extended to a basis contained in $X \cup Y$.*

(Proof omitted. Part (ii) follows by induction.)

(29.16) Lemma *Suppose K is a subfield of $\mathbf{C}$ and $u \in \mathbf{C}$ and $K[u]$ is a finite-dimensional vector space over K. Then u is algebraic over K, and its minimum polynomial has degree n.*

Proof. Let the dimension of $K[u]$ be n, so

$$1, u, u^2, \ldots, u^n$$

is not contained in any basis, so by the above lemma, is not linearly independent. Therefore there exists a list $\alpha_0, \ldots, \alpha_k$, where $k \leq n$ and $\alpha_k \neq 0$, such that

$$\alpha_0 + \alpha_1 u + \ldots \alpha_k u^k = 0.$$

Therefore u is algebraic over K, and its minimum polynomial has degree $\leq k$. By Lemma 29.12 (iv), $[K[u] : K] = k$, so $k = n$. Q.E.D.

(29.17) Lemma *Let $K \subseteq L \subseteq M$ be fields and suppose $[L : K] < \infty$ and $[M : L] < \infty$. Then $[M : K] = [M : L][L : K] < \infty$.*

Proof. Let $u_1, \ldots, u_k$ be a basis for L over K, and let $v_1, \ldots, v_m$ be a basis for M over L. Claim that

$$\{u_i v_j \colon 1 \leq i \leq k, 1 \leq j \leq m$$

is a basis for M as a vector space over K.

For any $x \in M$, $x = \sum \alpha_j v_j$ for unique scalars $\alpha_i \in L$. For each i, $\alpha_j = \sum \beta_{ij} u_i$ for unique scalars $\beta_{ij} \in K$. Then $x = \sum \beta_{ij} u_i v_j$ for unique scalars β_{ij}. Q.E.D.

(29.18) Corollary *If u is algebraic over K and v is algebraic over $K[u]$ then v is algebraic over K.*

Proof. $[K[v] : K] = [K[v] : K[u]][K[u] : K] < \infty$. Q.E.D.

30 Ruler-and-compass constructions

Ruler and compass constructions can be imagined to occur in the cartesian plane, where we begin, say, with the points $(0, 0$ and $(1, 0)$. The point $(0, 1)$ is easily constructed, and any point (x, y) where x and y are integers, or even rational.

In general, with ruler and compass, one can produce the line through two points, the point on a line closest to a given point, and

- The point where two non-parallel lines intersect
- The points where a line and a circle intersect
- The points where two circles intersect

(30.1) Definition *A real number u is* constructible *if the point $(u, 0)$ can be constructed in this way.*

(30.2) Proposition *A real number u is constructible if and only if there exists a list $a_1, \ldots, a_k$ of real numbers with $u \in \mathbf{Q}[a_1, \ldots, a_k]$ and $a_{i+1}^2 \in \mathbf{Q}[a_1, ldots, a_i]$ for $0 \leq i \leq k-1$ (Proof omitted.)*

(30.3) Corollary *If u is constructible, then it is algebraic and the degree of its minimum polynomial is a power of 2.*

(30.4) Corollary *The angle 20° cannot be constructed by ruler and compass.*

Proof. Let $x = 2\cos(20^\circ)$. It is enough to show that x is not a constructible real number. For any angle θ, by De Moivre's Theorem,

$$\cos(3\theta)+i\sin(3\theta) = (\cos(\theta)+i\sin(\theta))^3 = \cos^3(\theta)+3\cos^2(\theta)(i\sin(\theta))+3\cos(\theta)(i\sin(\theta))^2+(i\sin(\theta)^3,$$

$$\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta),$$

If $\theta = 20^\circ$ then $\cos(3\theta) = 1/2$, so

$$4(x/2)^3 - 3(x/2) - 1/2 = 0,$$

so $x^3 - 3x - 1 = 0$.

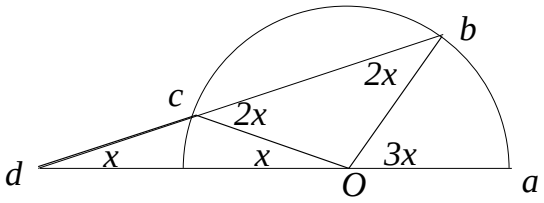
Replace x by $x + 1$, getting $x^3 + 3x^2 + 3$, This latter polynomial is irreducible by Eisenstein's Criterion, so $x^3 - 3x - 1$ is irreducible. Since the minimal polynomial of $2\cos(30^\circ)$ has degreee three, this number is not constructible, so $\cos(20^\circ)$ is not constructible. Q.E.D.

Since 60° is constrictible, the following is immmediate.

(30.5) Theorem *There is no general method to trisect an angle by ruler and compass.* ∎

There exist 'cheating' ruler-and-compass methods to trisect an angle, hinted at by the illustration below. To trisect the angle aOb, construct the semicircle as illustrated, with unit radius say. Take a ruler on which two points d and c have been marked, at unit distance apart. Keeping d on the baseline aO and c on the semicircle, adjust the placement of the ruler until it passes through the point b. Then $\widehat{adb} = (1/3)\widehat{aOb}$.

This goes beyond the 'legitimate' ruler-and-compass constructions which can solve only linear and quadratic equations.



31 Cubic equations

All polynomials $p(x)$ considered will have, say, rational coefficients: $p \in \mathbf{Q}[x]$, and we know that all their zeros can be found in the complex plane $\mathbf{C}$.

Remember that finding all complex solutions to an equation $p(x) = 0$ is equivalent to expressing it as a product of linear factors **(Factor Theorem).**

It is known (Fundamental Theorem of Algebra) that $\mathbf{C}$ is algebraically closed, that is, every polynomial $p(x)$ in $\mathbf{C}[x]$ of nonzero degree can be factored completely in $\mathbf{C}[x]$.

The familiar rule for factorising a polynomial $ax^2 + bx + c$ as $a(x - \alpha)(x - \beta)$ is

$$\alpha = (-b + \sqrt{b^2 - 4ac})/2a, \quad \beta = (-b - \sqrt{b^2 - 4ac})/2a$$

Notice that $\alpha + \beta = -b/a$ and $\alpha\beta = c/a$.

In $\mathbf{C}$ there are three distinct solutions to $x^3 - 1 = 0$. $x^3 - 1 = (x - 1)(x - \omega)(x - \omega^2)$ where $\omega = e^{2\pi i/3} = \cos(120^\circ + i\sin(120^\circ)$. Here $i = \sqrt{-1}$.

More simply, $\omega = (-1 + i\sqrt{3})/2$ and $\omega^2 = (-1 - i\sqrt{3})/2 = \overline{\omega} = 1/\omega$.

To solve the equation

$$ax^3 + bx^2 + cx + d = 0, \tag{31.1}$$

or equivalently, to factorise the polynomial, we apply several transformations. First

$$\text{divide by } a \text{ and substitute } x = y - \frac{b}{3a}$$

producing an equation for y in which the quadratic part vanishes: it has the form

$$y^3 + py + q = 0. \tag{31.2}$$

Next, substituting

$$y = z - \frac{p}{3z}$$

and multiplying by z^3 throughout produces an equation of the following form:

$$z^6 + qz^3 - \frac{p^3}{27} = 0 \tag{31.3}$$

which can be solved as a quadratic equation in z^3. From this all solutions to (31.1) can be found. Choose some α, and let $\beta = p/(3\alpha)$, so

$$z^6 + qz^3 - \frac{p^3}{27} = (z^3 - \alpha^3)(z^3 - \beta^3)$$

The six possible values of z are $\alpha, \omega\alpha, \omega^2\alpha, \beta, \omega\beta, \omega^2\beta$, where ω is a primitive cube root of unity, say $\omega = (-1+i\sqrt{3})/2 = e^{2\pi i/3}$. Then $z-p/(3z)$ takes on the values $\alpha+\beta$, $\omega\alpha+\omega^2\beta$, and $\omega^2\alpha+\omega\beta$. These are the three possible roots y, from which the three values of x can be deduced.

Example. $2x^3 + 3x^2 - 12x - 3 = 0.$ (31.4)

Divide by 2, and substitute $x = y - 1/2$.

$$y^3 - \frac{3}{2}y^2 + \frac{3}{4}y - \frac{1}{8} + \frac{3}{2}y^2 - \frac{3}{2}y + \frac{3}{8} - 6y + 3 - \frac{3}{2} = 0.$$

Simplify, getting the form (31.2).

$$y^3 - \frac{27}{4}y + \frac{7}{4} = 0.$$

Substitute $y = z + 9/4z$.

$$z^3 + \frac{27}{4}z + \frac{243}{16z} + \frac{729}{64z^3} - \frac{27}{4}z - \frac{243}{16z} + \frac{7}{4} = 0.$$

Multiply by z^3.

$$z^6 + \frac{7}{4}z^3 + \frac{729}{64} = 0.$$

So the form (31.3) has been reached; solve for z^3.

$$z^3 = \frac{-7 \pm i\sqrt{680}}{8}.$$

Let α be one of the six possible values of z, β its complex conjugate. The solutions for z then are

$$\alpha, \omega\alpha, \omega^2\alpha, \beta, \omega\beta, \omega^2\beta.$$

where $\omega = (-1 + i\sqrt{3})/2$ is a primitive cube root of 1. Its square is its conjugate: $\omega^2 = \bar{\omega}$.

Also

$$z^6 + \frac{7}{4}z^3 + \frac{729}{64} = (z - \alpha^3)(z - \beta^3).$$

Hence $(\alpha\beta)^3 = 729/64$. Since $\alpha\beta$ is positive, it must equal $9/4$, so $|\alpha| = 3/2$, and

$$\frac{9}{4\alpha} = \frac{9\bar{\alpha}}{4\alpha\bar{\alpha}} = \bar{\alpha} = \beta.$$

So with $z = \alpha$, $z + 9/(4z) = \alpha + \beta$. Similarly with $\omega\alpha$ and $\omega^2\alpha$. This gives us three, not six, solutions for y. Subtracting $1/2$ we get the solutions for the original equation.

To simplify this further, express α in polar form. Thus, $\alpha = |\alpha|e^{i\phi}$ where

$$\phi = \frac{\tan^{-1}(\frac{-\sqrt{680}}{7})}{3}.$$

Warning about inverse tan function. One must here be careful using $\tan^{-1}$. It has the range $(-\pi/2, \pi/2)$, whereas, in this example, 3ϕ is in the second quadrant (real part negative, imaginary part positive), between $\pi/2$ and π. $\tan(3\phi) = -3.72525848$, and applying inverse tan to this we get -1.30854116 radians. This is in the fourth quadrant. Now the tan function is periodic with period π, so we should add $\pi = 3.14159264$ to this number to get the correct value of 3ϕ: 1.83305148 radians. Therefore $\phi = .61101716$ radians.

Then $\alpha + \beta = 2r\cos(\phi) = 3\cos(\phi) = 2.45719467$. This is one of the solutions for y.

$$-\frac{1}{2} + 3\cos(\phi), -\frac{1}{2} + 3\cos(\phi + 2\pi/3), -\frac{1}{2} + 3\cos(\phi + 4\pi/3).$$

Using the simple UNIX 'bc' program gives the following numerical results: (In bc, variables are just one letter, so `p` means π and `f` means ϕ.)

```
  scale = 10
Calculate pi = 3.1415926532
  p=4*a(1)
Next, phi
  f = p-a(sqrt(680)/7)
  f=f/3
Now phi = .6110171617
  a=-.5+3*c(f)
First solution, a:
  1.9571946853
Check it:
  2*a*a*a+3*a*a=-12*a-3
  .0000000076
Second solution:
  f=f+2*p/3
  a=-.5+3*c(f)
  -3.2191159460
Check it:
  2*a*a*a+3*a*a-12*a-3
  .0000000284
Third solution:
  f=f+2*p/3
  a=-.5+3*c(f)
  -.2380787405
Check it:
  2*a*a*a+3*a*a-12*a-3
  .0000000321
```

Exercises. Solve the following.

(i) $2x^3 + 3x^2 - 12x + 7 = 0$ (ii) $2x^3 + 3x^2 - 12x + 17 = 0$.

32 The Galois group of an extension field

(32.1) Definition *Let $K \subseteq L$ be fields. The* Galois group *of L over K, $G(L/K)$, is the group of all those automorphisms of L which leave K fixed.*

(32.2) Example. $G(K/K) = \{e\}$, obviously.

(32.3) Example. $G(\mathbf{C}/\mathbf{R})$: let g be an automorphism of $\mathbf{C}$ which fixes $\mathbf{R}$.

Consider $g(i)$. Since $i^2 = -1$, $g(i)^2 = g(-1) = -1$ because g fixes -1. Therefore $g(i) = \pm(i)$.

For any complex number $x + iy$, $g(x + iy) = g(x) + g(i)g(y) = x + (g(i))y = x \pm iy$.

Thus g can be either the identity or complex conjugation. Since both of these are automorphisms which fix $\mathbf{R}$, $G(\mathbf{C}/\mathbf{R})$ is {identity, complex conjugation}.

(32.4) Definition *Suppose $g : L \to M$ is an isomorphism of fields. Sometimes to avoid clutter we write ga rather than $g(a)$. Given a polynomial $p(x) \in L(x)$, suppose*

$$p(x) = a_0 + a_1x + \ldots + a_nx^n.$$

Then $g(p)$ or gp is the polynomial

$$ga_0 + ga_1x + \ldots + ga_nx^n.$$

(32.5) Lemma *With $g\colon L \to M$ as above, for any $u \in L$, $g(p(u)) = (gp)(gu)$. (Trivial).* ∎

In studying $G(\mathbf{C}/\mathbf{R})$, the observation that $g(i)^2 = -1$ was important. It is generalised by the following observation: **elements of $G(L/K)$ permute the roots of polynomials**. This is put more directly in the following corollary to the above lemma.

(32.6) Corollary *Let $g \in G(L/K)$. For any $u \in L$, if u is algebraic over K with minimum polynomial p, then $g(u)$ is algebraic with the same minimum polynomial.*

Proof. Since $g \in G(L/K)$, $g(p) = p$. Therefore $p(gu) = (gp)(gu) = g(p(u)) = g0 = 0$. Q.E.D.

(32.7) Example. Let $K + \mathbf{Q}$, $L = K[\sqrt{2}]$. The minimum polynomial of $\sqrt{2}$ over K is $x^2 - 2$. Elements of the Galois group must permute the roots of this polynomial. The candiates are $x+y\sqrt{2} \mapsto x \pm y\sqrt{2}$. Both of these are automorphisms.

(32.8) Example. Let $K + \mathbf{Q}$, $L = K[\sqrt[3]{2}]$. The minimum polynomial of $\sqrt[3]{2}$ over K is $x^3 - 2$. Any element of the Galois group must permute these roots. However, L contains only one of the three roots, so $G(L/K) = \{e\}$.

(32.9) Example. Suppose $\omega = e^{2pi/3}$. a cube root of 1, and $u = \sqrt[3]{2}$. Let $K + \mathbf{Q}[\omega]$, $L = K[u]$.

The minimum polynomial of ω and ω^2 is $x^2 + x + 1$. As is always the case with quadratic extension fields. $G(K/\mathbf{Q})$ has two elements, the identity and that sending ω to ω^2 (actually just complex conjugation).

The minimum polynomial of u over $\mathbf{Q}$ is $x^3 - 2$.

Observe that $x^3 - 2$ is the minimum polynomial of u *over* K. $[\mathbf{Q}[u] : \mathbf{Q}] = 3$, whereas $[K : \mathbf{Q}] = 2$. Therefore, by the degree theorem for extension fields. $u \notin K$. The same goes for the other roots of the polynomial, namely ωu and $\omega^2 u$. Since K contains no root of $x^3 - 2$, it has no linear factor over K, so (being cubic) $x^3 - 2$ is irreducible over K, as claimed.

Therefore $[L : K] = 3$. This time L contains all three roots of the minimum polynomial of u over K. L has a basis $1, u, u^2$ as a vector space over K. Any automorphism of L which fixes K is defined uniquely by its effect on u, and it must send u to $u, \omega u,$ or $\omega^2 u$. The first possiblity gives the identity automorphiism. Consider the second, where $u \mapsto \omega u$.

For any $a, b, c \in K$ (the coefficients are complex numbers), $a + bu + cu^2 \mapsto a + b\omega u + c\omega^2 u^2$. This is clearly additive and clearly bijective. To check the multiplicative property, note that $u^3 = 2$ and $u^4 = 2u$, so

$$(a + bu + cu^2)(c + du + eu^2) = ac + 2be + 2cd + (ad + 2ce)u + bdu^2$$

and

$$(a + b\omega u + c\omega^2 u^2)(c + d\omega u + e\omega^2 u^2) = ac + 2be + 2cd + (ad + 2ce)\omega u + bd\omega^2 u^2$$

so the map is multiplicative, an automorphism. Similarly there is a unique automorphism taking u to $\omega^2 u$. Thus $G(L/K)$ is cyclic of order 3.

Also, conjugation is an automorphism of L fixing $\mathbf{Q}$, so $G(L/\mathbf{Q})$ contains an automorphism exchanging ωu with $\omega^2 u$. This is an element of order 2, and $G(L/\mathbf{Q}) \supseteq G(L.K)$ contains a 3-cycle on the three root $u, \omega u, \omega^2 u$. Thus $G(L/\mathbf{Q})$ permutes these roots in all possible ways, and every permutation defines a unique automorphism, so $G(L/\mathbf{Q}) \cong S_3$.

(32.10) Definition *Let $J \subseteq N$ be fixed fields. For any field L with $J \subseteq L \subseteq N$, let L' be that subgroup of automorphisms $\alpha \in G(N/J)$ which fix L.*

The fields J and N are left implicit in this notation.

All extensions are assumed to have finite degree. Note that if $L \subseteq M$ then $M' \subseteq L'$.

(32.11) Lemma *Let J and N be fixed (with $[N : J] < \infty$). For any K, L, M with $J \subseteq K \subseteq M \subseteq N$, $[K' : M'] \leq [M : K]$.*

Proof. By induction. Obviously true if $M = K$. If there exists a field L with $K \subset L \subset M$ and K, L, M distinct, then one can assume by induction that $[L' : M'] \leq [M : L]$ and $[K' : L'] \leq [L : K]$, so $[K' : M'] = [K' : L'][L' : M'] \leq [L : K][M : L] = [M : K]$. Hence we should assume that there is no field properly between K and M, in which case $M = K[u]$ for some u algebraic over K.

Observe that for any left coset of $K[u]'$, the effect on u is the same. In other words, if $g_1, g_2 \in K'$ where $g_1^{-1}g_2 \in K[u]'$, $g_1^{-1}g_2(u) = u$ so $g_1 u = g_2 u$. On the other hand, if $g_1, g_2 \in K'$ and $g_1 u = g_2 u$, then $g_1^{-1}g_2 u = u$, and $g_1^{-1}g_2$ fixes K, so $g_1^{-1}g_2$ fixes $K[u]$. Therefore $[K' : K[u]']$ is the number of different values of gu, where g runs over K'. Since gu must always be a root of the same irreducible polynomial (over K), there are $[K[u] : K]$ possibilities. Hence $[K' : M'] \leq [M : K]$ as required. Q.E.D.

(32.12) Corollary *If $K \subseteq M$ are fields and $[M : K] \leq \infty$ then $|G(M/K)| \leq [M : K]$.*

Proof. With $K = J$ and $M = N$, $K' = G(N/J) = G(M/K)$ and $M' = \{e\}$ so $[K' : M'] = |G(M/K)| \leq [M : K]$. Q.E.D.

33 Normal field extensions

(33.1) Definition *Let $K \subseteq M$ be fields. Then M is a* normal extension *of K if for every $u \in M \backslash K$ there exists $g \in G(M/K)$ such that $gu \neq u$.*

As usual, we are considering only subfields of $\mathbf{C}$.

Furthermore, we are considering only extensions of finite degree of $\mathbf{Q}$. These contain only algebraic numbers.

(33.2) Lemma *Let M be a normal extension of K and let u be any element of M. Since it is assumed that M has finite degree, u is algebraic over K. Let $p(x)$ be its minimum polynomial over K. Then M contains* all *the roots of $p(x)$.*

Proof. Let $u_1, \ldots, u_m$ be the roots of p in M, so $m \geq 1$. Write $p(x) = (x-u_1)\cdots(x-u_m)q(x)$, so $q(x) = p/((x-u_1)\cdots(x-u_m))$. Since every element of $G(M/K)$ only permutes the roots of p in M, $gq = q$ for every such g. Thus all the coefficients of q are fixed by g. Since M is normal over K, this implies that all coefficents if q are in K, so $q \in K[x]$ is a proper divisor of p. Since p is irreducible, q is constant and all the roots of p are in M. Q.E.D.

(33.3) Lemma *Let M be a normal extension of K, with $[M : K] < \infty$. Then $|G(M/K)| = [M : K]$.*

Sketch of proof. This argument requires some linear algebra, and we only give a sketch.

Suppose that $u_1, \ldots, u_m$ are a basis for M over K, and suppose $G(M/K) = g_1, \ldots, g_k$, where $k < m = [M : K]$. The $k \times m$ matrix $g_i u_j$ has coefficients in M.

In the usual way, define the kernel of this matrix as the set of all lists $a_1, \ldots, a_m$ of elements of M such that $\sum g_i u_j a_j = 0$, $1 \leq i \leq k$.

Since it has fewer rows than columns, the kernel contains at least one such list $a_1, \ldots, a_m \in M$, wher enot all are zero.

Without loss of generality the first ℓ numbers a_j are nonzero, and the last $m - \ell$ are zero, $a_1 = 1$, and the list has been chosen to make ℓ as small as possible (as many elements as possible are zero).

For any $g \in G(M/K)$ and $1 \leq i \leq k$, $g^{-1}g_i$ is another element of $G(M/K)$, so, $\sum g^{-1}g_i u_j a_j = 0$. Applying g, we get $\sum g_i u_j (g a_j) = 0$. Thus the list $g a_j$ is also in the kernel.

One of the equations is given by the identity automorphism, so one of the equations says that $\sum u_j a_j = 0$. Since the u_j are linearly independent over K, it follows that not all the a_j are in K. Since M is a normal extension, there exists a g in $G(M/K)$ and an a_j such that $g a_j \neq a_j$. On the other hand, $g a_1 = g1 = 1 = a_1$, Therefore the list $b_1 = a_1 - g a_1, b_2 = a_2 - g a_2, \ldots, b_m = a_m - g a_m$ is not all zero but has fewer nonzero elements than $a_1, a_2, \ldots, a_m$. On the other hand, it is the difference of two elements of the kernel, so it is also in the kernel, and has fewer nonzero elements, a contradiction. Hence $k \geq m$.

But we already know that $|G(M/K)| \leq [M : K]$ (Corollary 32.12), so $|G(M/K) = [M : K]$. Q.E.D.

(33.4) Lemma *If $|G(M/K)| = [M : K] < \infty$ then M is a normal extension.*

Proof. Let L be the subfield of M fixed by all automorphisms in $G(M/K)$. Note $G(M/L) = G(M/K)$, and M is a normal extension of L. Also $[M : K] = |G(M/K)| = |G(M/L)| \leq [M : L]$. Since $K \subseteq L$, $L = K$. Q.E.D.

34 Splitting fields and stable intermediate fields

(34.1) Definition *Given fields $K \subseteq L \subseteq M$, L is a* stable intermediate field *if for all $g \in G(M/K)$,, $g(L) \subseteq L$.*

(34.2) Lemma *If $K \subseteq L \subseteq M$, M is a normal extension of K, and $[M : K] < \infty$, and L is a stable intermediate field, let $G = G(M/K)$ and H be the subgroup of all automorphisms in G which fix L. Let ρ be the 'restriction map' which sends every $g \in G$ to the restriction of g to L.*

Then L is a normal extension of K, M is a normal extension of L, ρ is a homomorphism, its image is $G(L/K)$, and its kernel is H, so $H \lhd G(M/K)$ and $G(L/K) \cong G/H$.

Proof. Since L is stable, for every $g \in G$ the restriction $\rho(g)$ maps L into L, so $\rho(g) \in G(L/K)$.

If $u \in L \backslash K$, then $u \in M \backslash K$, so there exists a $g \in G$ such that $gu \neq u$. Then $(\rho(g))u = gu \neq u$, so L is a normal extension.

Given $g_1, g_2 \in G$, $x \in L$,

$$(g_1 \circ g_2)x = g_1(g_2 x) = g_1(\rho(g_2)x) = (\rho g_1)((\rho g_2)(x)) = \rho(g_1) \circ \rho(g_2)(x),$$

so ρ is a homomorphism into $G(L/K)$.

For any $g \in G$, g fixes L if and only if $\rho(g)$ is the identity in $G(L/K)$, so the kernel of ρ is H, which is therefore a normal subbgroup of G, and the image of ρ is isomorphic to G/H.

Finally, $|\rho(G)| = |G|/|H| \geq [M : K]/[M : L] = [L : K]$, so $\rho(G) = G(L/K)$ and L is a normal extension of K. Then $|H| = [M : L] = [M : K]/[L : K]$ so M is a normal extension of L. Q.E.D.

(34.3) Definition *Let $\mathbf{Q} \subseteq K \subseteq L \subseteq \mathbf{C}$ be fields. If there exists a polynomial f in $K[x]$ such that L is generated by all the roots of f, then L is called a* splitting field *over K. If K is not mentioned then it is assumed that $K = \mathbf{Q}$.*

(34.4) Definition *Let M be a normal extension of K, $K \subseteq L \subseteq M$ a splitting field over K. Then L is a stable intermediate field.*

Proof. Suppose f is the polynomial whose roots generate L over K. Then any element of $G(M/K)$ merely permutes the roots of f, so L is a stable intermediate field. Q.E.D.

35 Radical field extensions and solvability

(35.1) Definition *A* radical extension *of a field L is either one of the form $L[u]$ where the minimum polynomial of u over L is $x^n - a$ for some integer n and $a \in L$, or (recursively) a radical extension of a radical extension.*

A polynomial equation $p(x) = 0$ is solvable by radicals *if there exists a radical extension of $\mathbf{Q}$ containing at least one root of p.*

(35.2) Lemma *Let L be a (finite-degree) normal extension of K and $u \in \mathbf{C}$ has minimum polynomial $x^p - a$ over L where p is prime. Suppose also that L contains all p-th roots of unity. Then $L[u]$ is a normal extension of K, and if $|G(L/K)| = [L : K]$ then $G(L[u]/K)| = [L]u] : K]$..*

Proof.

36 Appendix: the natural number system

The natural numbers, or nonnegative integers, are denoted $\mathbf{N}$. Their general properties are summed up in the following axioms, called *Peano's postulates.*

- 0 is a natural number.
- If x is a natural number, then so its its successor $x + 1$.
- 0 is not a successor, i.e. for no natural number x is $x + 1 = 0$.
- Cancellation: if $x + 1 = y + 1$ then $x = y$.
- Principle of induction: if $P(x)$ is a property of natural numbers x such that $P(0)$ and $P(x) \Rightarrow P(x + 1)$ for all natural numbers x, then $P(x)$ is true for all natural numbers x.

Addition and multiplicaion of natural numbers x is expressed by some equations:

- $x + 0 = x$
- $x + *y + 1) = (x + y) + 1$
- $x0 = 0$
- $x(y + 1) = xy + x$

There are two other principles which are related to, or equivalent to, induction. That is, **Course of values induction:**

If $P(x)$ is a property of natural numbers such tthat for every x, $((\forall y < x)P(y)) \Rightarrow P(x)$, then $P(x)$ holds for every x.

(36.1) Also, the principle of well-ordering (PWO): every nonempty subset of $\mathbf{N}$ has a smallest element.

37 Appendix: cardinality