RINGS, FIELDS, AND VECTOR SPACES

S. K. Jain *Ohio University*

I.	Axioms, Examples, and Types of Rings	209
II.	Ideals	212
III.	Homomorphisms	213
IV.	Unique Factorization and Euclidean Domains	214
V.	Rings and Fields of Fractions	215
VI.	Vector Spaces	215
VII.	Algebraic Extensions of a Field	216
VIII.	Normal and Separable Extensions	218
IX.	Finite Fields	219
X.	Fundamental Theorem of Galois Theory and Application	219
XI.	Ruler and Compass Constructions	220

GLOSSARY

Algebraically closed field: Field having no proper algebraic extensions.

Algebraic extension: Extension of a field such that each element of the extension satisfies a nontrivial polynomial over the base field.

Basis (vecto space): Set of linearly independent elements generating the space.

Dimension (vector space): Number of elements in any basis.

Division ring: Ring with unity in which each nonzero element is invertible.

Homomorphism: Function from a ring to a ring that preserves binary operations.

Ideal: A subring of a ring that is closed under multiplication by elements of the ring and those of the subring.

Integral domain: Ring in which the product of nonzero elements is not zero.

Isomorphism: Homomorphism that is both 1–1 and onto.

Normal extension: Extension of a field such that whenever it contains one root of an irreducible polynomial over the base field, then it must contain all the roots.

Splitting field (of a polynomial): Smallest extension containing the roots of the polynomial.

A ring is an algebraic structure with two binary operations written additively and multiplicatively such that it is an abelian group under addition, it is a semigroup under multiplication, and the multiplication is distributive over addition. For example, the set of integers $\mathbb{Z}$ and the set of rational numbers $\mathbb{Q}$ are rings. If each nonzero element of a ring has multiplicative inverse and multiplication is commutative, then such a ring is called a field, e.g., $\mathbb{Q}$ is a field. A vector space is an additive abelian group whose elements, known as vectors, can be suitably multiplied by the elements, known as scalars, from some field; and the multiplication of vectors with scalars obeys natural laws, such as $(\mathbf{x} + \mathbf{y})\alpha = \mathbf{x}\alpha + \mathbf{y}\alpha$, $\mathbf{x}(\alpha + \beta) = \mathbf{x}\alpha + \mathbf{x}\beta$, where $\mathbf{x}$, $\mathbf{y}$ are vectors and α, β are scalars. The ordinary geometric vectors, that is, directed line segments, form a vector space over the field of real numbers.

I. Axioms, Examples, and Types of Rings

The term "ring" originated in the study of arithmetic properties of integers and algebraic integers (see Definition VII.6). Hilbert studied the sets of the form $\mathbb{Z}[\alpha] = \{a + b\alpha \mid a, b \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of all integers and α is a root of $x^2 + px + q$, $p, q \in \mathbb{Z}$ and called such a set a Zahlring (number ring). It was A. Franklin in 1914 who made a general study of the abstract structure underlying such sets $\mathbb{Z}[\alpha]$ and defined the term ring (Definition I.3) of which $\mathbb{Z}[\alpha]$ is a concrete example.

It is essentially believed that the study of the theory of rings and ideals was initially inspired by attempts to solve Fermat's last problem by several mathematicians, including Lamé (1795–1870), Kummer (1810–1893), Dirichlet (1805–

ENCYCLOPEDIA OF PHYSICAL SCIENCE
AND TECHNOLOGY, VOL. 12

1859), and Dedekind (1831–1916). Later, the problems in algebraic geometry gave further impetus to the subject. [*See* ALGEBRAIC GEOMETRY.]

Whereas rings are algebraic structures with two binary operations usually written additively and multiplicatively satisfying certain laws. (Definition I.3), semigroups and groups are algebraic structures with one binary operation.

I.1 Definition. If $(\mathbf{S}, .)$ is a nonempty set with a multiplicative binary operation such that

S1 (Associative law): $a(bc) = (ab)c$

for all $a, b, c \in \mathbf{S}$, then $(\mathbf{S}, .)$ or simply $\mathbf{S}$ is called a (multiplicative) semigroup.

I.2 Definition. Suppose $(\mathbf{G}, .)$ is a semigroup satisfying the following axioms:

G1 (Existence of identity): $\exists e \in \mathbf{G}$ such that $ae = a$, $\forall a \in \mathbf{G}$

G2 (Existence of inverse): $\forall a \in \mathbf{G}$, $\exists b \in \mathbf{G}$ such that $ab = e$.

Then $(\mathbf{G}, .)$ or simply $\mathbf{G}$ is called a (multiplicative) group.

It follows, as a consequence, that if $\mathbf{G}$ is a group, then the element e in axiom G1 is unique and so is the element b in axiom G2. Further, it follows that $ae = a = ea$ for all $a \in \mathbf{G}$, and $ab = e = ba$. The element e is called the identity element and is denoted by 1 (read as one); the element b in G2 is called the inverse of a and is denoted by a^{-1}. If $\mathbf{G}$ is a group with binary operation written additively (i.e., $a + b$), then the identity element is denoted by 0 (read as zero) and the inverse of a is denoted by $-a$.

If $(\mathbf{G}, .)$ is a group satisfying the condition

$$ab = ba \qquad \forall a, b \in \mathbf{G}$$

then $\mathbf{G}$ is called an abelian group. For an additive abelian group this condition shall read as $a + b = b + a$.

I.3 Definition. Let $(\mathbf{R}, +, .)$ be a nonempty set with two binary operations $+$ and $.$ satisfying the following axioms:

R1 $(\mathbf{R}, +)$ is an additive abelian group;
R2 $(\mathbf{R}, \cdot)$ is a multiplicative semigroup;
R3 (Left distributive law): $a(b + c) = ab + ac$, (Right distributive law): $(a + b)c = ac + bc$;
for all $a, b, c \in R$.

Then $(\mathbf{R}, +, \cdot)$ or simply $\mathbf{R}$ is called a ring.

A ring $\mathbf{R}$ is called commutative if $ab = ba$ for all $a, b \in \mathbf{R}$. If a ring is not commutative, it is called noncommutative.

A ring $\mathbf{R}$ is said to have unity or identity element if there exists $e \in \mathbf{R}$ such that $ae = a = ea$ for all $a \in \mathbf{R}$. The element e is generally denoted by 1.

It follows from the ring axioms R1–R3 that if $\mathbf{R}$ is a ring, then, as one would expect,

(i) $a0 = 0 = 0a$
(ii) $a(-b) = -(ab) = (-a)b$
(iii) $a(b - c) = ab - ac$, $(a - b)c = ac - bc$, for all $a, b, c \in \mathbf{R}$.

Now, $a0 = a(0 + 0) = a0 + a0$ and so $a0 + (-a0) = a0$ or $0 = a0$. Similarly, $0a = 0$ and other assertions.

Further, generalized distributive laws also hold, that is, for all $a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_n \in \mathbf{R}$,

(iv) $(a_1 + a_2 + \cdots + a_m)(b_1 + b_2 + \cdots + b_n) = a_1b_1 + a_1b_2 + \cdots + a_1b_n + a_2b_1 + a_2b_2 + \cdots + a_2b_n + \cdots + a_mb_1 + a_mb_2 + \cdots + a_mb_n$

Next, if $a \in \mathbf{R}$, then as a matter of notation

(v) $a^m = \overbrace{a\,a\cdots a}^{m \text{ times}}$ (m, a positive integer)

(vi) $ma = \overbrace{a + a + \cdots + a}^{m \text{ times}}$ (m, a positive integer)

(vii) $ma = \overbrace{(-a) + (-a) + \cdots + (-a)}^{-m \text{ times}}$ (m, a negative integer)

(viii) $a^{-m} = \overbrace{a^{-1} \cdot a^{-1} \cdot \cdots \cdot a^{-1}}^{m \text{ times}}$, if a has an inverse a^{-1} (m, a positive integer)

(ix) $a^0 = 1$, if $1 \in R$.
(x) $0a = 0$, where 0 on the left is the number zero and 0 on the right is the zero of the ring.

I.4 Examples. $\mathbb{Z}$: the ring of integers under usual addition and multiplication; $\mathbb{Q}$: the ring of rational numbers under usual addition and multiplication; $\mathbb{R}$: the ring of real numbers under

usual addition and multiplication; $\mathbb{C}$: the ring of complex numbers under usual addition and multiplication; and $\mathbb{Z}/(n)$, $\mathbb{Z}_n$: the ring of integers modulo n, where n is a positive integer. Let

$$\bar{a} = \{\ldots, a - 2n, a - n, a, a + n, a + 2n, \ldots\},$$
$$a \in \mathbb{Z}$$

Define $\bar{a} = \bar{b}$ if $a - b$ is a multiple of n. Let $\mathbb{Z}/(n) = \{\bar{a} | a \in \mathbb{Z}\}$. Define addition and multiplication in $\mathbb{Z}/(n)$ as follows:

$$\bar{a} + \bar{b} = \overline{a + b}$$
$$\bar{a}\bar{b} = \overline{ab}$$

Then the set $\mathbb{Z}/(n)$ is a ring under binary operations defined above. The members of $\mathbb{Z}/(n)$ are equivalence classes determined by the relation "$\equiv$" on $\mathbb{Z}$ defined by $a \equiv b$ if $n|(a - b)$. $\mathbb{Z}/(n)$ is also denoted by $\mathbb{Z}_n$. $\mathbb{R}_n$: the ring of $n \times n$ matrices over $\mathbb{R}$ under usual addition and multiplication of matrices, that is, if

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix},$$

and

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1n} \\ \vdots & & \vdots & & \vdots \\ b_{i1} & \cdots & b_{ij} & \cdots & b_{in} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nn} \end{bmatrix}$$

then

$$B = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1j} + b_{1j} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} + b_{i1} & \cdots & a_{ij} + b_{ij} & \cdots & a_{in} + b_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nj} + b_{nj} & \cdots & a_{nn} + b_{nn} \end{bmatrix}$$

and

$$AB = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}b_{k1} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kj} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kn} \\ \vdots & & \vdots & & \vdots \\ \sum_{k=1}^{n} a_{ik}b_{k1} & \cdots & \sum_{k=1}^{n} a_{ik}b_{kj} & \cdots & \sum_{k=1}^{n} a_{ik}b_{kn} \\ \vdots & & \vdots & & \vdots \\ \sum_{k=1}^{n} a_{nk}b_{k1} & \cdots & \sum_{k=1}^{n} a_{nk}b_{kj} & \cdots & \sum_{k=1}^{n} a_{nk}b_{kn} \end{bmatrix}$$

If $\mathbf{R}$ is any ring, then the set of $n \times n$ matrices with entries from $\mathbf{R}$ can similarly be made into a ring $\mathbf{R}_n$, called the ring of matrices over $\mathbf{R}$. $\mathbb{R}[x]$: the ring of polynomials $a_0 + a_1x + \cdots + a_nx^n$ in a variable x with coefficients $a_i \in \mathbb{R}$, where n is any nonnegative integer, under the usual addition and multiplication of two polynomials, namely,

$$(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_mx^m) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots,$$

and

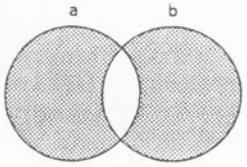
$$(a_0 + a_1x + \cdots + a_nx^n)(b_2 + b_2x + \cdots + b_mx^m) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \cdots$$

Similarly, if $\mathbf{R}$ is any ring, the polynomials $a_0 + a_1x + \cdots + a_nx^n$, $a_i \in \mathbf{R}$, $n \geq 0$, form a ring $\mathbf{R}[x]$.

$\mathbf{P}(A)$: the ring of all subsets of a set A under addition and multiplication defined as follows:

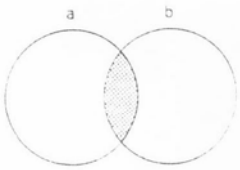
$$a + b = (a \cup b) - (a \cap b)$$

i.e., $a + b$ is the shaded area



$$ab = a \cap b$$

i.e., ab is the shaded area



where $a, b \in \mathsf{P}(A)$. $\mathsf{P}(A)$ is a commutative ring with identity (which is the whole set A). The zero element is the empty set. This ring has a property that $x^2 = x$ and $2x = 0$ for all $x \in \mathsf{R}(A)$. A ring R with the property $x^2 = x$ for all $x \in \mathsf{R}$ is called a Boolean ring. Such rings have found some very interesting applications in electrical engineering and computer science.

I.5 Definition. A ring R whose nonzero elements form a group under multiplication is called a division ring. If, in addition, R is commutative, then R is called a field.

Dedekind introduced the term Zahlkörper (= number field) to denote a set of complex numbers satisfying the field axioms. The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are examples of fields. Also, the set $\{\alpha \in \mathbb{C} | \alpha$ is a root of $a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]\}$, called the set of algebraic numbers, forms a field. It was E. Steinitz who first gave the abstract definition of a field in 1910. In his important paper "The Algebraic Theory of Fields," he took up the task of investigating the common properties underlying several concrete examples constructed previously from the subsets of complex numbers. A field can be finite too (consider, e.g., $\mathbb{Z}(p)$, p prime). Wedderburn (in 1905) showed that every finite division ring must be a field.

An example of a division ring that is not a field was first discovered by W. Hamilton (1843) as a certain subset of the ring of 2×2 matrices over $\mathbb{C}$, now called the ring of real quaternions.

I.6 Ring of Real Quaternions. Let $Q = \{\left[\begin{smallmatrix} a & b \\ -\bar{b} & \bar{a} \end{smallmatrix}\right] \mid a, b \in \mathbb{C}\}$, where a, b denote the complex conjugates of a, b, respectively. It is easy to see that if $A = \left[\begin{smallmatrix} a & b \\ -\bar{b} & \bar{a} \end{smallmatrix}\right]$ is a nonzero element of Q, then $\det A$, the determinant of A, is $a\bar{a} + b\bar{b} \neq 0$, and so A^{-1} exists. Q is not commutative, and hence Q is a division ring. Systems such as Q, matrices over $\mathbb{C}$, were known earlier under the name hypercomplex systems.

I.7 Definition. A ring R is called an integral domain if $ab = 0$, $a, b \in \mathsf{R}$ implies $a = 0$ or $b = 0$.

The ring of integers is an integral domain. Every field and thus every division ring is an integral domain. If $n > 1$, the $n \times n$ matrices over any field is not an integral domain. For let $a = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $b = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Then $a \neq 0$, $b \neq 0$ but $ab = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right] = 0$, the zero matrix.

I.8 Definition. If there exists a positive integer n such that $na = 0$ for each $a \in \mathsf{R}$, the smallest such positive integer is called the characteristic of R. If no such positive integer exists, R is said to have characteristic zero (or infinity according to some authors).

Every integral domain has characteristic either 0 or prime. The characteristic of $\mathbb{Z}/(n)$, the ring of integers modulo n, is n.

I.9 Definition. Let $\mathsf{R}_1, \cdots, \mathsf{R}_n$ be a family of rings. Then the cartesian product $\mathsf{R} = \mathsf{R}_1 \times \cdots \times \mathsf{R}_n$ can be made into a ring by defining pointwise addition and multiplication. This ring R is called the direct product of $\mathsf{R}_1, \ldots, \mathsf{R}_n$.

If (R_i), $i = 1, 2, 3, \ldots$, is an infinite family of rings, then one can similarly define the direct product ΠR_i of the family (R_i), $i = 1, 2, 3, \ldots$. If S is a subring of $\Pi \mathsf{R}_i$ such that each element of S is a sequence with finitely many nonzero terms, then S is called the direct sum of the family (R_i), $i = 1, 2, 3, \ldots$, and is denoted by $\bigoplus \Sigma \mathsf{R}_i$.

II. Ideals

Two distinguished subsets of a ring are subrings and ideals (introduced by Kummer and Dedekind in connection with Fermat's last problem). Subrings, as one would expect, is a subset S of a ring R such that S itself is a ring under the same binary operations as on R. Any subring S, like any subgroup in a group, does not always yield a canonical quotient ring. This is remedied by introducing the notion of an ideal (analogous to the normal subgroup in a group).

II.1 Definition. A nonempty subset S of a ring R is called a right (left) ideal if

(i) $a - b \in \mathbf{S}$ for all $a, b \in \mathbf{S}$, and
(ii) $ar \in \mathbf{S}$ ($ra \in \mathbf{S}$) for all $a \in \mathbf{S}$, $r \in \mathbf{R}$.

A right ideal or a left ideal is a subring. But, in general, not every subring is a right or left ideal. For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$ but $\mathbb{Z}$ is not a right or left ideal in $\mathbb{Q}$, $\mathbb{Z}$ and $\mathbb{Z}/(n)$ are the only rings with unity having the property that each subring is an ideal.

II.2 Definition. A nonempty subset of a ring that is both a right ideal and a left ideal is called an ideal (also called two-sided ideal).

If $\mathbf{R}$ is commutative, every right or left ideal is an ideal. (0) and $\mathbf{R}$ are ideals in any ring $\mathbf{R}$, called trivial ideals. If $\mathbf{A}$ is a right (or left) ideal in a ring $\mathbf{R}$ such that $\mathbf{A}$ contains an invertible element a, then $\mathbf{A} = \mathbf{R}$. For $aa^{-1} = 1 = a^{-1}a$ implies $1 \in \mathbf{A}$ and so $r \in \mathbf{A}$ for all $r \in \mathbf{R}$, since $\mathbf{A}$ is a right (or left) ideal. Thus $\mathbf{A} = \mathbf{R}$. In particular, a field has no ideals other than (0) and $\mathbf{R}$. Indeed, if $\mathbf{R}$ is commutative ring with more than one element and if $\mathbf{R}$ has no nontrivial ideals, then $\mathbf{R}$ is a field.

II.3 Definition. A ring with no nontrivial ideals is called a simple ring.

As remarked above, every commutative simple ring with more than one element is a field. An example of a noncommutative simple ring is the ring $\mathbf{F}_n$ of $n \times n$ matrices over a field $\mathbf{F}$, $n > 1$.

II.4 Definition. If $\mathbf{R}$ is a ring with unity, the right (or left) ideal $a\mathbf{R}$ (or $\mathbf{R}a$) is called a principal right (or left) ideal. A ring is called a principal right (or left) ideal ring if each right (or left) ideal is principal.

II.5 Definition. If $\mathbf{R}$ is a commutative ring, then principal right (or left) ideal ring is called principal ideal ring. If, in addition, $\mathbf{R}$ is an integral domain, then $\mathbf{R}$ is called a principal ideal domain, usually called PID.

Examples of PID are $\mathbb{Z}$ and $\mathbf{F}[x]$. The polynomial ring $\mathbf{D}[x]$ over a commutative integral domain $\mathbf{D}$ is a PID iff $\mathbf{D}$ is a field.

II.6 Definition. Let $\mathbf{R}$ be a ring and $\mathbf{I}$ be a two-sided ideal. Let $a, b \in \mathbf{R}$. Define a relation "$\equiv$" on $\mathbf{R}$ by $a \equiv b$ if $a - b \in \mathbf{I}$. Then "$\equiv$" is an equivalence relation (i.e., this is reflexive, symmetric, and transitive). Let $\bar{a}$ denote the equivalence class containing a. Then $\bar{a} = \{a + x | x \in I\}$. The set $\mathbf{R}/\mathbf{I}$ (also written as $\bar{\mathbf{R}}$) of equivalence classes can be made into a ring by defining $\bar{a} + \bar{b} = \overline{a + b}$, $\bar{a}\bar{b} = \overline{ab}$. This ring $\mathbf{R}/\mathbf{I}$ is called the quotient ring of $\mathbf{R}$ modulo $\mathbf{I}$.

The zero element $\bar{0}$ of $\mathbf{R}/\mathbf{I}$ is $\bar{a}$ for any $a \in \mathbf{I}$. If $1 \in \mathbf{R}$ then $\bar{1}$ is the unity element of $\mathbf{R}/\mathbf{I}$. If $\mathbf{R}$ is commutative, then $\mathbf{R}/\mathbf{I}$ is also commutative. The right, left, or two-sided ideals of $\mathbf{R}/\mathbf{I}$ are of the form $\mathbf{A}/I$, where $\mathbf{A}$ is a right, left, or two-sided ideal of $\mathbf{R}$ containing $\mathbf{I}$.

II.7 Definition. An ideal $\mathbf{M}$ in a ring $\mathbf{R}$ is called maximal if $\mathbf{M} \neq \mathbf{R}$, and whenever $\mathbf{M}$ is properly contained in an ideal $\mathbf{N}$ then $\mathbf{N} = \mathbf{R}$.

It thus follows that $\mathbf{M}$ is a maximal ideal iff $\mathbf{R}/\mathbf{M}$ is a simple ring. Thus, if $\mathbf{R}$ is a commutative ring with unity, then $\mathbf{M}$ is a maximal ideal in $\mathbf{R}$ iff $\mathbf{R}/\mathbf{M}$ is a field. In particular, if $\mathbf{R} = \mathbf{F}[x]$ is a polynomial ring over a field $\mathbf{F}$ and $p(x)$ is an irreducible polynomial over $\mathbf{F}$ (i.e., $p(x) \notin \mathbf{F}$, and if $p(x) = p_1(x)p_2(x)$, $p_1(x), p_2(x) \in \mathbf{F}[x]$ then $p_1(x)$ or $p_2(x) \in \mathbf{F}$) then $(\mathbf{F}[x])/(p(x))$ is a field. This field contains a carbon copy of $\mathbf{F}$ (usually identified with F itself) and is called an extension of $\mathbf{F}$. The field of complex numbers may be looked upon as the field $(\mathbb{R}[x])/(x^2 + 1)$ whose elements are of the form $\{a + b\bar{x} | a, b \in \mathbb{R}\}$, where $\bar{x}^2 = -1$ ($\bar{x}$ is usually written as $\sqrt{-1}$ or i).

II.8 Definition. Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ be a family of right (left, or two-sided) ideals of a ring $\mathbf{R}$. Then the smallest right (left, or two-sided) ideal containing each $\mathbf{A}_i$ is called the sum of $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ written as $\mathbf{A}_1 + \cdots + \mathbf{A}_n$ or $\sum_{i=1}^n \mathbf{A}_i$.

The sum of right ideals $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ is easily seen to be the set $\{a_1 + a_2 + \cdots + a_n | a_i \in \mathbf{A}_i\}$.

II.9 Definition. The sum $\mathbf{A} = \sum_{i=1}^n \mathbf{A}_i$ of right (left, or two-sided) ideals is called direct if each element a of $\mathbf{A}$ can be uniquely expressed as $a = a_1 + a_2 + \cdots + a_n$, $a_i \in \mathbf{A}$. If the sum $\mathbf{A} = \sum_{i=1}^n \mathbf{A}_i$ is direct, it is written as $\mathbf{A} = \bigoplus \sum_{i=1}^n \mathbf{A}_i$.

The following is a useful set of equivalent statements for a sum to be direct:

(1) $\mathbf{A} = \sum_{i=1}^n \mathbf{A}_i$ is a direct sum.
(2) If $0 = \sum_{i=1}^n a_i$, $a_i \in \mathbf{A}_i$, then $a_i = 0$, $1 \leq i \leq n$.
(3) $\mathbf{A}_i \cap \sum_{j=1, j\neq i}^n \mathbf{A}_j = (0)$, $1 \leq i \leq n$.

III. Homomorphisms

Homomorphisms are mappings between algebraic structures preserving binary operations.

III.1 Definition. Let $\mathbf{R}, \mathbf{S}$ be rings and f: $\mathbf{R} \to \mathbf{S}$ be a mapping such that $f(a + b) = f(a) - f(b)$ and $f(ab) = f(a)f(b)$, for all $a, b \in \mathbf{R}$. Then

f is called a ring homomorphism or simply homomorphism from R to S.

If f is onto, then $\mathbf{S}$ is called a homomorphic image of $\mathbf{R}$ (under f). If f is 1-1, then $\mathbf{R}$ is said to be embeddable in $\mathbf{S}$ (equivalently, $\mathbf{S}$ is said to contain a carbon copy of $\mathbf{R}$): in this case, $\mathbf{R}$ is called isomorphic into $\mathbf{S}$. If f is both 1-1 and onto, then f is called an isomorphism from $\mathbf{R}$ onto $\mathbf{S}$, and the ring $\mathbf{R}$ is said to be isomorphic onto the ring $\mathbf{S}$, written as $\mathbf{R} \simeq \mathbf{S}$. Being "isomorphic onto" is an equivalence relation in the class of rings. From an abstract point of view, isomorphic rings may be regarded as the same ring.

III.2 Definition. Let f: $\mathbf{R} \to \mathbf{S}$ be a homomorphism from a ring $\mathbf{R}$ to a ring $\mathbf{S}$. The sets $\operatorname{Ker} f = \{a \in \mathbf{R} | f(a) = 0\}$, $\operatorname{Im} f = \{f(a)\, a \in \mathbf{R}\}$ are called the kernel and the image of f respectively.

$\operatorname{Ker} f$ is an ideal of $\mathbf{R}$ but $\operatorname{Im} f$ is a subring of $\mathbf{S}$. Further, $\operatorname{Ker} f = (0)$ iff f is 1–1, and $\operatorname{Im} f = \mathbf{S}$ iff f is onto. The fundamental theorem of homomorphism states that if f: $\mathbf{R} \to \mathbf{S}$ is a homomorphism, then $\mathbf{R}/\operatorname{Ker} f \simeq \operatorname{Im} f$. The other important theorem about homomorphisms is the so-called correspondence theorem: Let f: $\mathbf{R} \to \mathbf{S}$ be a homomorphism of a ring $\mathbf{R}$ onto a ring $\mathbf{S}$. Then $A \to f(A)$ defines a 1–1 order-preserving correspondence of the set of all right, left, or two-sided ideals of $\mathbf{R}$ that contain $\operatorname{Ker} f$ onto the set of all right, left, or two-sided ideals of $\mathbf{S}$.

If $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ is a family of ideals in a ring $\mathbf{R}$ such that $\mathbf{A}_i + \mathbf{A}_j = \mathbf{R}$, $i \neq j$, and $1 \leq i, j \leq n$, then $\mathbf{R}/\cap_{i=1}^{n} \mathbf{A}_i \simeq \mathbf{R}/\mathbf{A}_1 \times \mathbf{R}/\mathbf{A}_2 \times \cdots \times \mathbf{R}/\mathbf{A}_n$. In particular, if $a_1, a_2, \ldots, a_n \in \mathbf{R}$, then there exists $a \in \mathbf{R}$ such that $a \equiv a_i \pmod{A_i}$ for all i. This is known as the Chinese Remainder Theorem. Applying the above to the ring of integers $\mathbb{Z}$ and ideals $(p_1^{e_1}), \ldots, (p_n^{e_n})$, where p_i are distinct primes, $e_i \geq 0$, one obtains $\mathbb{Z}/(m) \simeq \mathbb{Z}/(p_1^{e_1}) \times \cdots \times \mathbb{Z}/(p_n^{e_n})$, where $m = p_1^{e_1} \cdots p_n^{e_n}$. The nontrivial ideals of $\mathbb{Z}/(p_i^{e_i})$ are $(p_i)/(p_i^{e_i})$, $(p_i^2)/(p_i^{e_i})$, $\ldots$, $(p_i^{e_i-1})/(p_i^{e_i})$ and hence these are linearly ordered, that is, given any two ideals $\mathbf{A}$, $\mathbf{B}$, either $\mathbf{A} \subset \mathbf{B}$ or $\mathbf{B} \subset \mathbf{A}$.

IV. Unique Factorization and Euclidean Domains

Let $\mathbf{R}$ be a commutative integral domain with unity. Such a ring is also simply known as a domain. If a, b are nonzero elements in $\mathbf{R}$, we say that b divides a (or b is a divisor of a) and that a is divisible by b (or a is a multiple of b) if there exists in $\mathbf{R}$ an element c such that $a = bc$. This is written as $b|a$ or $a \equiv 0 \pmod{b}$. Thus, an element u in $\mathbf{R}$ is a unit, that is, invertible iff u is a divisor of 1. Two elements $a, b \in \mathbf{R}$ are called associates if there exists a unit $u \in \mathbf{R}$ such that $a = bu$. This means that a and b are associates iff $a|b$ and $b|a$. An element $b \in \mathbf{R}$ is called an improper divisor of $a \in \mathbf{R}$ if b is either a unit or an associate of a. A nonzero element $a \in \mathbf{R}$ is called an irreducible element if (i) a is not a unit, and (ii) every divisor of a is improper. A nonzero element $p \in \mathbf{R}$ is called a prime if (i) p is not a unit, and (ii) if $p|ab$, $a, b \in \mathbf{R}$ then $p|a$ or $p|b$. A prime element is always irreducible but not conversely, for $2 + \sqrt{5}$ is irreducible but is not prime in the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} | a, b \in \mathbb{Z}\}$. However, an irreducible element in a commutative principal ideal domain is always prime.

IV.1 Definition. A commutative integral domain with unity is called a unique factorization domain (or briefly, a UFD) if it satisfies the following conditions: (i) every nonunit of $\mathbf{R}$ is a finite product of irreducible factors, and (ii) every irreducible element is prime. Indeed, it follows then that the factorization into irreducible elements is unique.

Any principal ideal domain is a UFD. In particular, the ring of integers $\mathbb{Z}$ and the ring of polynomials $\mathbf{F}[x]$ over a field $\mathbf{F}$ are both unique factorization domains. The ring $\mathbb{Z}[\sqrt{-5}]$ is not a UFD, since $9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5})$ and each of the elements $3, 2 + \sqrt{-5}, 2 - \sqrt{-5}$ is irreducible in $\mathbb{Z}[\sqrt{-5}]$. Further, if $\mathbf{R}$ is a UFD then the polynomial ring $\mathbf{R}[x]$ is also a UFD. This yields, in particular, that $\mathbb{Z}[x]$ is a UFD. Note, however, $\mathbb{Z}[x]$ is not a PID, since the ideal $(2, x)$ is not principal.

In a unique factorization domain $\mathbf{R}$, one may define a greatest common divisor (GCD) of a pair of elements $a, b \in \mathbf{R}$ as an element $d \in \mathbf{R}$ such that (i) $d|a$ and $d|b$, and (ii) if $c|a$ and $c|b$ then $c|d$, $c \in \mathbf{R}$. If $\mathbf{R}$ is a UFD, then there exists a GCD of any pair of elements that is uniquely determined to within unit factors, denoted by (a, b). Thus, (a, b) is a set in which any two elements are associates. We write $(a, b) = c$ to mean that (a, b) consists of all unit multiples of c. Two elements a, b in a UFD are called relatively prime if $(a, b) = 1$. The following are some interesting properties of the greatest common divisor: (1) $c(a, b) = (ca, cb)$; (2) if $(a, b) = 1$ and $b|ac$, then $b|c$; (3) if $(a, b) = 1$ and if $a|c$ and $b|c$, then $ab|c$; and (4) if $(a, b) = 1$ and $(a, c) = 1$, then $(a, bc) = 1$.

An important example of unique factorization domain is a Euclidean domain, that is, a domain admitting division algorithm.

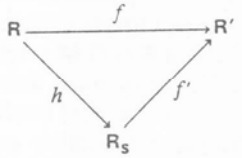
IV.2 Definition. A commutative integral domain with unity is called Euclidean domain if there exists a function ϕ: $E \to \mathbb{Z}$ satisfying the following axioms: (i) If $a, b \in E^* = E - \{0\}$ and $b|a$, then $\phi(b) \leq \phi(a)$, (ii) For each pair of elements $a, b \in E$, $b \neq 0$, there exists elements q and r in E such that $a = bq + r$ with $\phi(r) < \phi(b)$.

IV.3 Examples. $\mathbb{Z}$ where ϕ is given by $\phi(n) = |n|$ (i.e., the absolute value of n); $\mathbf{F}[x]$, $\mathbf{F}$ a field, where ϕ is given by $\phi(f(x)) =$ degree of $f(x)$ if $f(x) \neq 0$ and $\phi(0) = -1$; and $\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} | a, b \in \mathbb{Z}\}$, where ϕ is given by $\phi(m + n\sqrt{-1}) = m^2 + n^2$.

Every Euclidean domain is a PID. For if $\mathbf{A}$ is a nonzero ideal in a Euclidean domain $\mathbf{R}$, $\mathbf{A} = (d)$ where $\phi(d)$ is the smallest integer in the set $\{\phi(a) | 0 \neq a \in \mathbf{A}\}$. Not every PID is a Euclidean domain. The domain $\mathbb{Z}[\sqrt{-19}] = \{a + b\sqrt{-19} | a, b \in \mathbb{Z}$ and a, b are both odd or both even$\}$ is a PID but not a Euclidean domain.

V. Rings and Fields of Fractions

Let $\mathbf{R}$ be a commutative ring with unity and $\mathbf{S}$ a nonempty subset of $\mathbf{R}$ such that $ab \in \mathbf{S}$ for all $a, b \in \mathbf{S}$. Define a relation $\sim$ on $\mathbf{R} \times \mathbf{S}$ by $(a, s) \sim (a', s')$ iff $\exists s'' \in \mathbf{S}$ such that $s''(as' - a's) = 0$. "$\sim$" is an equivalence relation. Let R_S denote the set of equivalence classes. Denote by a/s the equivalence class containing (a, s). $\mathbf{R_S}$ can be made into a ring by defining $a_1/s_1 + a_2/s_2 = a_1s_2 + a_2s_1/s_1s_2$ and $a_1/s_1 \cdot a_2/s_2 = a_1a_2/s_1s_2$. The ring $\mathbf{R_S}$ is called the ring of fractions with respect to a subset $\mathbf{S}$, or the localization of $\mathbf{R}$ at $\mathbf{S}$. Suppose $1 \in \mathbf{S}$. Then there is a canonical homomorphism h: $\mathbf{R} \to \mathbf{R_S}$ given by $h(a) = a/1$ such that (i) Ker $h = \{a \in \mathbf{R} | as = 0$ for some $s \in S\}$; (ii) the elements of $h(\mathbf{S})$ are invertible in $\mathbf{R_S}$; (iii) every element of $\mathbf{R_S}$ can be written as $h(a)/h(s)$, $a \in \mathbf{R}$, $s \in \mathbf{S}$; (iv) if f: $\mathbf{R} \to \mathbf{R}'$ is any homomorphism of a ring $\mathbf{R}$ into a ring $\mathbf{R}'$ such that every element $f(s)$, $s \in \mathbf{S}$, is invertible in $\mathbf{R}'$, then there exists a homomorphism f': $\mathbf{R_S} \to \mathbf{R}'$ such that the following diagram is commutative



that is, $f'h = f$. This is also expressed by saying that any such homomorphism f factors through h.

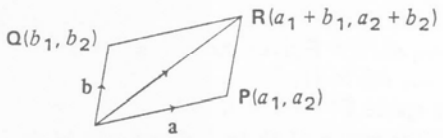
By taking $\mathbf{R}$ to be a commutative integral domain, and $\mathbf{S} = \mathbf{R} - \{0\}$, one obtains that $\mathbf{R}$ is embeddable in a field $\mathbf{R_S}$, called the field of fractions. Another interesting case occurs by taking $\mathbf{S} = \mathbf{R} - \mathbf{P}$, where $\mathbf{P}$ is any prime ideal (i.e., $ab \in \mathbf{P}$, $a, b \in \mathbf{R} \Rightarrow a \in \mathbf{P}$ or $b \in \mathbf{P}$). In this case $\mathbf{R_S}$ (also written as $\mathbf{R_P}$) has a unique maximal ideal given by $\{a/s | a \in \mathbf{P}, s \in \mathbf{S}\}$. Such a ring (i.e., a ring having a unique maximal ideal) is called a local ring. The local ring $\mathbf{R_P}(= \mathbf{R_S})$ is called the localization at the prime ideal $\mathbf{P}$.

Malcev (1937) gave an example of an integral domain that is not embeddable in a division ring. A necessary and sufficient condition that $\mathbf{R}$ be embeddable in a division ring is $a\mathbf{R} \cap b\mathbf{R} \neq 0$ for all $0 \neq a, b \in \mathbf{R}$ (or $\mathbf{R}a \cap \mathbf{R}b \neq 0$ for all $0 \neq a, b \in \mathbf{R}$). These conditions, discovered by Öre, are known as right Öre-condition (or left Öre-condition). For a commutative integral domain these conditions are clearly satisfied.

VI. Vector Spaces

Directed line segments in a plane or space are used to represent velocity, force, etc. Such a physical quantity is thus represented by a line segment OP, of the appropriate length and direction, drawn from the origin 0 of coordinates. If the end point P has coordinates (a_1, a_2), the vector OP is completely determined by these coordinates and we may write vector $OP = (a_1, a_2)$. Any ordered pair of real numbers a_1, a_2 defines a vector in a plane in this way.

If $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ are two vectors in a plane, then the sum of $\mathbf{a}$ and $\mathbf{b}$ is formed by adding the components, that is, $\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2)$. This corresponds to the parallelogram law



Furthermore, on multiplying $\mathbf{a}$ by a real number α, one obtains another vector $\alpha\mathbf{a} = (\alpha a_1, \alpha a_2)$. These geometric interpretations have motivated the study of abstract vector spaces, which is the most applicable branch of abstract algebra studied under the title "linear algebra."

VI.1 Definition. Let $\mathbf{V}$ be an additive abelian group, $\mathbf{F}$ a field, and suppose for each $\alpha \in \mathbf{F}$, $x \in \mathbf{V}$ there is a unique element $\alpha x \in \mathbf{V}$ satisfying the following axioms:

V1 $\quad \alpha(x + y) = \alpha x + \alpha y$

V2 $\quad (\alpha + \beta)x = \alpha x + \beta x$

V3 $(\alpha\beta)x = \alpha(\beta x)$

V4 $1x = x$

for all $\alpha, \beta \in \mathbf{F}$ and $x, y \in \mathbf{V}$. Then $\mathbf{V}$ is called a vector space over $\mathbf{F}$. The elements of $\mathbf{V}$ are called vectors and the elements of $\mathbf{F}$ are called scalars. αx is generally called the scalar multiplication of α with x.

VI.2 Examples. In the following examples $\mathbf{F}$ denotes any field.

$\mathbf{F}^n$: the vector space of n-tuples $(a_1, a_2, \ldots, a_n)$, $a_i \in \mathbf{F}$ where addition and scalar multiplication are pointwise. For $n = 2$ and 3, one obtains the usual vector space of directed line segments in a plane and space, respectively.

$\mathbf{F}[x]$: the vector space of polynomials in x over $\mathbf{F}$ where addition and scalar multiplication are the usual operations on polynomials.

$\mathbf{F}^{m\times n}$: the vector space of $m \times n$ matrices with entries from $\mathbf{F}$, where addition and scalar multiplication are the usual operations on matrices. In particular, the vector spaces $\mathbf{F}^{m\times 1}$ and $\mathbf{F}^{1\times n}$ are generally denoted by $\mathbf{F}^m$ and $\mathbf{F}^n$, respectively, without any ambiguity as the context always makes it clear whether the elements are written as column vectors or as row vectors.

$\mathbf{V}^S$: the vector space over $\mathbf{F}$ consisting of all mappings from a set S to a vector space $\mathbf{V}$ over $\mathbf{F}$ where addition and scalar multiplication are usual operations on mappings, that is, given f: $\mathbf{S} \to \mathbf{V}$, g: $\mathbf{S} \to \mathbf{V}$, and $\alpha \in \mathbf{F}$, $s \in S$, define $f + g$, αf: $S \to \mathbf{V}$ by $(f + g)(s) = f(s) + g(s)$, $(\alpha f)(s) = \alpha f(s)$.

From the vector space $\mathbf{V}^S$, one obtains a large family of vector spaces. For example, $S = \{1, 2, \ldots, n\}$, $\mathbf{V} = \mathbf{F}$ gives the vector space $\mathbf{F}^n$; $S = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$, $\mathbf{V} = \mathbf{F}$, gives the vector space $\mathbf{F}^{m\times n}$; $S = [a, b]$ and $\mathbf{V} = \mathbb{R}$ gives the vector space of all real-valued functions on the closed interval $[a, b]$, and so on.

VI. 3 Definition. Let $\mathbf{V}$ be a vector space over a field $\mathbf{F}$. A subset $\mathbf{W}$ of $\mathbf{V}$ is called a subspace of $\mathbf{V}$ if $\mathbf{W}$ is itself a vector space over $\mathbf{F}$ under the same operations, that is, addition and scalar multiplication, as in $\mathbf{V}$.

It follows that a nonempty subset $\mathbf{W}$ of a vector space $\mathbf{V}$ over $\mathbf{F}$ is a subspace iff for all $x, y \in \mathbf{W}$ and $\alpha \in \mathbf{F}$, we have $x - y \in \mathbf{W}$ and $\alpha x \in \mathbf{W}$.

Let $\mathbf{V}$ be a vector space over a field $\mathbf{F}$ and $\mathbf{S}$ be a nonempty subset of $\mathbf{V}$. Then any element of $\mathbf{V}$ of the form $\alpha_1 x_1 + \cdots + \alpha_m x_m$, $x_i \in \mathbf{S}$, $\alpha_i \in \mathbf{F}$, is called a linear combination of elements of $\mathbf{S}$. A subset $\mathbf{S}$ is said to be a linearly independent set if for every finite sequence of distinct elements $x_1, \ldots, x_m$ of $\mathbf{S}$, $\alpha_1 x_1 + \cdots + \alpha_m x_m = 0$, $\alpha_i \in \mathbf{F}$, implies each $\alpha_i = 0$. A subset $\mathbf{S}$ of $\mathbf{V}$ is called linearly dependent if $\mathbf{S}$ is not linearly independent, that is, there exists a finite sequence of distinct elements $x_1, \ldots, x_m$ of $\mathbf{S}$ such that $\alpha_1 x_1 + \cdots + \alpha_m x_m = 0$, $\alpha_i \in \mathbf{F}$, and at least some $\alpha_i \neq 0$.

A vector space $\mathbf{V}$ over a field $\mathbf{F}$ is said to be generated by a subset $\mathbf{S}$ of $\mathbf{V}$ if each element of $\mathbf{V}$ is a linear combination of a finite number of elements in $\mathbf{S}$. For example, the vector space $\mathbf{F}[x]$ is generated by $\mathbf{S} = \{1, x, x^2, x^3, \cdots\}$. Another example is the subset $\mathbf{S} = \{(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1^{\text{th}}, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)\}$, which generates $\mathbf{F}^n$. Although the whole set $\mathbf{V}$ always generates $\mathbf{V}$, the interest lies in some kind of a "minimal" subset.

VI.4 Definition. A subset $\mathbf{S}$ of a vector space $\mathbf{V}$ is called a basis if $\mathbf{S}$ generates $\mathbf{V}$ and $\mathbf{S}$ is linearly independent.

The first fundamental result in the theory of vector spaces is that every vector space has a basis. Furthermore, if $\mathbf{S}$ and $\mathbf{S}'$ are two bases of a vector space $\mathbf{V}$, then $|\mathbf{S}| = |\mathbf{S}'|$, that is, they have the same cardinality. If $\mathbf{S}$ is a basis of $\mathbf{V}$, then the cardinality of $\mathbf{S}$ is called the dimension of $\mathbf{V}$, written as dim $\mathbf{V}$. A vector space $\mathbf{V}$ is said to be finite or infinite dimensional according as dim $\mathbf{V}$ is finite or infinite. If a vector space $\mathbf{V}$ is generated by a finite subset $\mathbf{S}$, then there exists a linearly independent subset $\mathbf{X}$ of $\mathbf{S}$ that generates $\mathbf{V}$.

VII. Algebraic Extensions of a Field

The theory of fields is the richest and perhaps the most useful branch of algebra. Its usefulness lies in providing analytic tools for problems in geometry, number theory, coding theory, and its richness in the profundity and variety of results obtained.

Let $\mathbf{F}$ be a field and $\mathbf{F}[x]$ the ring of polynomials in x over $\mathbf{F}$. $\mathbf{F}[x]$ is a principal ideal domain and hence a unique factorization domain. Thus, $\mathbf{M}$ is a maximal ideal in $\mathbf{F}[x]$ iff $\mathbf{M} = (p(x))$ where $p(x)$ is an irreducible polynomial, that is, $p(x) = f(x)g(x)$ where $f(x), g(x) \in \mathbf{F}[x]$ implies $f(x)$ or $g(x) \in \mathbf{F}$. Irreducible polynomials over a field play an important role in field theory. Let $f(x) \in \mathbf{F}[x]$ be a polynomial over $\mathbf{F}$ and let $\mathbf{E}$ be a field containing $\mathbf{F}$ as a subfield. $\alpha \in \mathbf{E}$ is called a zero or root of $f(x) = a_0 + a_1 x + \cdots + a_n \alpha x^n$ if $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$. This is expressed by saying

$f(\alpha) = 0$. By division algorithm, it follows that if α is a root of $f(x)$, then $x - \alpha$ is a factor of $f(x)$. Thus, a polynomial $f(x) \in \mathbf{F}[x]$ having a root $\alpha \in \mathbf{F}$ must be reducible over $\mathbf{F}$ (i.e., not irreducible). However, if $f(x) \in \mathbf{F}[x]$ is reducible, it is not necessary that $f(x)$ has a root in $\mathbf{F}$. [Consider, e.g., $(x^2 + 1)(x^2 + 1) \in \mathbb{R}[x]$.] In a special case, if $f(x) \in \mathbf{F}[x]$ is a polynomial of degree at most 3, then $f(x)$ is reducible over $\mathbf{F}$ iff $f(x)$ has a root in $\mathbf{F}$. An interesting result for reducibility of polynomials over $\mathbb{Q}$ is Gauss's lemma, which states that $f(x) \in \mathbb{Q}[x]$ is reducible over $\mathbb{Q}$ iff it is reducible over the ring $\mathbb{Z}$. A criterion for testing a polynomial $f(x) \in \mathbb{Z}[x]$ to be irreducible is due to Eisenstein. The Eisenstein criterion states that if $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$, and if there exists a prime p such that $p \mid a_0, \ldots, p \mid a_{n-1}$, $p \nmid a_n$, $p^2 \nmid a_0$, then $f(x)$ is irreducible over $\mathbb{Z}$, and hence over $\mathbb{Q}$. In particular, $x^4 + 5x^3 + 10x^2 + 10x + 5 \in \mathbf{Z}[x]$ is irreducible over $\mathbb{Q}$.

If $\mathbf{F}$ is a subfield of $\mathbf{E}$, then $\mathbf{E}$ is called an extension field of $\mathbf{F}$, or simply an extension of $\mathbf{F}$. If $\mathbf{E}$ is an extension of $\mathbf{F}$, then trivially $\mathbf{E}$ is a vector space over $\mathbf{F}$. The dimension of $\mathbf{E}$ over $\mathbf{F}$ is usually written as $[\mathbf{E} : \mathbf{F}]$, called the degree of $\mathbf{E}$ over $\mathbf{F}$. If $[\mathbf{E} : \mathbf{F}] < \infty$, then $\mathbf{E}$ is called a finite extension, else $\mathbf{E}$ is called an infinite extension of $\mathbf{F}$. Suppose there is a nonzero homomorphism of a field $\mathbf{F}$ to $\mathbf{E}$. Choose a set $\mathbf{S}$ disjoint from $\mathbf{F}$ and having the same cardinality as that of $\mathbf{E} - \mathbf{F}$. Let $\mathbf{K} = \mathbf{F} \cup \mathbf{S}$. Then $\mathbf{K}$ can be made into a field by defining addition and multiplication in an obvious way so that $\mathbf{F}$ is a subfield of $\mathbf{K}$ and $\mathbf{K} \simeq \mathbf{E}$. Thus, $\mathbf{K}$ is an extension of $\mathbf{F}$. Identifying $\mathbf{F}$ with its copy in $\mathbf{E}$, it is common to say that $\mathbf{E}$ is an extension of $\mathbf{F}$. Given an irreducible polynomial $p(x)$ over a field $\mathbf{F}$, the field $\mathbf{E} = (\mathbf{F}[x])/(p[x])$ contains a carbon copy of $\mathbf{F}$ (under the embedding $a \to \bar{a} = a + (p(x))$, $a \in \mathbf{F}$) and so $\mathbf{E}$ is an extension of $\mathbf{F}$. Identifying $\mathbf{F}$ with its image in $\mathbf{E}$, it then follows that the polynomial $p(x) = a_0 + a_1x + \cdots + a_nx^n$ has a root $\bar{x} = x + (p(x))$ in $\mathbf{E}$. Therefore, every irreducible polynomial (hence, every polynomial) over $\mathbf{F}$ has a root in some extension of $\mathbf{F}$, a result usually known as Kronecker's theorem.

VII.1 Definition. Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$. An element $\alpha \in \mathbf{E}$ is said to be algebraic over $\mathbf{F}$ if there exist elements $a_0, a_1, \ldots, a_n$ ($n \geq 1$) of $\mathbf{F}$, not all zero, such that $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$. In other words, an element $\alpha \in \mathbf{F}$ is algebraic over $\mathbf{F}$ if there exists a nonconstant polynomial $p(x) \in \mathbf{F}[x]$ such that $p(\alpha) = 0$.

The following is a basic result for an algebraic element. Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$ and $\alpha \in \mathbf{E}$ be algebraic over $\mathbf{F}$. Let $p(x) \in \mathbf{F}[x]$ be a polynomial of least degree such that $p(\alpha) = 0$. Then (i) $p(x)$ is irreducible over $\mathbf{F}$; (ii) if $g(x) \in \mathbf{F}[x]$ is such that $g(\alpha) = 0$, then $p(x) \mid g(x)$; (iii) there is exactly one monic polynomial $m(x) \in \mathbf{F}[x]$ of least degree such that $m(\alpha) = 0$; and (iv) $\mathbf{F}[x]/(p(x)) \simeq \mathbf{F}[\alpha] = \mathbf{F}(\alpha)$ and $[\mathbf{F}[\alpha] : \mathbf{F}] = n$, where n is the degree of $p(x)$. [In other words, if α is algebaic over $\mathbf{F}$, the polynomials in α over $\mathbf{F}$ forms a field and hence coincides with the smallest subfield $\mathbf{F}(\alpha)$ of $\mathbf{E}$ containing α and $\mathbf{F}$.] The unique monic polynomial $m(x) \in \mathbf{F}[x]$ of least degree satisfied by α is called the minimal polynomial of α over $\mathbf{F}$.

VII.2 Definition. Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$. If each element of $\mathbf{E}$ is algebraic over $\mathbf{F}$, then $\mathbf{E}$ is called an algebraic extension of $\mathbf{F}$.

If $\mathbf{E}$ is a finite extension of $\mathbf{F}$ and $\alpha \in \mathbf{E}$, then $1, \alpha, \ldots, \alpha^n$ must be linearly dependent over $\mathbf{F}$, where $[\mathbf{E} : \mathbf{F}] = n$. Thus, $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$, $a_i \in \mathbf{F}$, and not all a_i are zero. Therefore, α is algebraic over $\mathbf{F}$, and hence $\mathbf{E}$ is algebraic extension of $\mathbf{F}$. But not every algebraic extension must be finite extension. For example, $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \ldots, \sqrt{p}, \ldots)$, where p is prime, is an algebraic extension of $\mathbb{Q}$, but it is not a finite extension.

Let $\mathbf{E}$ be an extension of $\mathbf{F}$ and $a, b \in \mathbf{E}$ be algebraic over $\mathbf{F}$. Then $[\mathbf{F}(a) : \mathbf{F}] < \infty$, and regarding b to be algebraic over $\mathbf{F}(a)$, $[\mathbf{F}(a, b) : \mathbf{F}(a)] < \infty$, thus $[\mathbf{F}(a, b) : \mathbf{F}] < \infty$. In particular, $a \pm b$, ab, a/b (if $b \neq 0$) are all algebraic over $\mathbf{F}$. Therefore, the subset of $\mathbf{E}$ consisting of all algebraic elements over $\mathbf{F}$ forms a subfield of $\mathbf{E}$ called the algebraic closure of $\mathbf{F}$ in $\mathbf{E}$. A useful fact about an algebraic extension $\mathbf{E}$ of $\mathbf{F}$ is that any embedding of $\mathbf{E}$ into itself over $\mathbf{F}$ (i.e., which keeps each element of $\mathbf{F}$ fixed) is an automorphism. (A 1–1 homomorphism of a field onto itself is called an automorphism of the field.)

By Kronecker's theorem every polynomial $f(x)$ over a field $\mathbf{F}$ has a root in some extension $\mathbf{E}$ of $\mathbf{F}$. This implies that there exists an extension $\mathbf{K}$ of $\mathbf{F}$ that contains all the roots of $f(x)$.

VII.3 Definition. If K is an extension of $\mathbf{F}$ such that $[\mathbf{K} : \mathbf{F}]$ is least with the property that $\mathbf{K}$ contains all the roots of $f(x) \in \mathbf{F}[x]$, then $\mathbf{K}$ is called a splitting field of $f(x)$ over $\mathbf{F}$. Splitting field of a polynomial is unique up to isomorphism.

VII.4 Definition. A field $\mathbf{K}$ is called algebraically closed if it possesses no proper algebraic extensions.

For any field $\mathbf{K}$ the following are equivalent: (i) $\mathbf{K}$ is algebraically closed, (ii) every irreducible polynomial over $\mathbf{K}$ is of degree 1, (iii) every nonconstant polynomial over $\mathbf{K}$ factors into linear factors in $\mathbf{K}[x]$, and (iv) every nonconstant polynomial has at least one root in $\mathbf{K}$.

VII.5 Definition. Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$. $\mathbf{E}$ is called an algebraic closure of $\mathbf{F}$ if (i) $\mathbf{E}$ is an algebraic extension of $\mathbf{F}$, and (ii) $\mathbf{E}$ is algebraically closed.

It requires technical details to prove the existence and uniqueness (upto isomorphism) of algebraic closure of a field. The algebraic closure of a field $\mathbf{F}$ is generally denoted by $\bar{\mathbf{F}}$. The algebraic closure of $\mathbb{R}$ is $\mathbb{C}$. The elements of the algebraic closure $\bar{\mathbb{Q}}$ are called algebraic numbers, that is, $\alpha \in \mathbb{C}$ is an algebraic number if α is algebraic over $\mathbb{Q}$.

VII.6 Definition. An element $\alpha \in \mathbb{C}$ is called an algebraic number (algebraic integer) if α is a root of a polynomial (a monic polynomial) over $\mathbb{Z}$.

VIII. Normal and Separable Extensions

Let $(f_i(x))_{i\in\wedge}$ be a family of polynomials of degree ≥ 1 over a field $\mathbf{F}$. By a splitting field of a family $(f_i(x))_{i\in\wedge}$ of polynomial over $\mathbf{F}$ is meant an extension $\mathbf{E}$ of $\mathbf{F}$ such that all the polynomials $f_i(x)$ split into linear factors in $\mathbf{E}[x]$, and $\mathbf{E}$ is generated over $\mathbf{F}$ by their roots. Also, $\mathbf{E}$ is unique (upto isomorphism). The following are equivalent statements: (i) $\mathbf{E}$ is splitting field of a family of polynomials over $\mathbf{F}$, (ii) every irreducible polynomial in $\mathbf{F}[x]$ that has a root in $\mathbf{E}$ splits into linear factors in $\mathbf{E}[x]$, and (iii) every embedding of $\mathbf{E}$ into $\bar{\mathbf{F}}$ that keeps each element of $\mathbf{F}$ fixed is an automorphism of $\mathbf{E}$.

VIII.1 Definition. An extension $\mathbf{E}$ of a field $\mathbf{F}$ is called normal extension of $\mathbf{F}$ if $\mathbf{E}$ satisfies any one of the equivalent statements mentioned above.

$\mathbb{C}$ is a normal extension of $\mathbb{R}$, but $\mathbb{R}$ is not a normal extension of $\mathbb{Q}$ since there exists an irreducible polynomial $x^3 + 2 \in \mathbb{Q}[x]$ that has one root in $\mathbb{R}$ but does not split into linear factors in $\mathbb{R}[x]$.

VIII.2 Definition. An irreducible polynomial $p(x)$ in $\mathbf{F}[x]$ is called a separable polynomial if all its roots are simple, that is, each root is of multiplicity one. Any polynomial $f(x) \in \mathbf{F}[x]$ is called separable if all its irreducible factors are separable. A polynomial that is not separable is called inseparable.

VIII.3 Definition. Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$. An algebraic element $\alpha \in \mathbf{E}$ is called separable if the minimal polynomial of α over $\mathbf{F}$ is separable. An algebraic extension $\mathbf{E}$ of a field $\mathbf{F}$ is called separable extension if each element of $\mathbf{E}$ is separable over $\mathbf{F}$. An algebraic extension $\mathbf{E}$ of a field $\mathbf{F}$ is called inseparable if it is not separable.

Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbf{F}[x]$. Define $f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1}$. An element α in $\mathbf{E}$, an extension of $\mathbf{F}$, is a multiple root iff $f'(\alpha) = 0$. In particular, if $f(x)$ is an irreducible polynomial over $\mathbf{F}$, then $f(x)$ has a multiple root iff $f'(x) = 0$. However, if the characteristic of $\mathbf{F}$ is zero, $f'(x) = 0$ implies $a_1 = a_2 = \cdots = a_n = 0$, yielding that $f(x) = a_0$ a contradiction, since $f(x)$ is irreducible. Therefore, each irreducible polynomial over a field of characteristic zero has distinct roots. Consequently, any algebraic extension of a field of characteristic zero is separable.

Let $\mathbf{K} = \mathbf{F}(x)$ be the field of rational functions in x over a field $\mathbf{F}$ of characteristic 3. Then the polynomial $y^3 - x \in \mathbf{K}[y]$ is an irreducible polynomial and has all its roots equal, say, each being α. Hence, $\mathbf{K}(\alpha)$ is an inseparable extension of K.

VIII.4 Definition. An extension $\mathbf{E}$ of a field $\mathbf{F}$ is called simple extension if $\mathbf{E} = \mathbf{F}(\alpha)$ for some $\alpha \in \mathbf{E}$.

A necessary and sufficient condition that a finite extension $\mathbf{E}$ of a field $\mathbf{F}$ is simple is that there are only a finite number of intermediate fields between $\mathbf{F}$ and $\mathbf{E}$. A useful and interesting result is that a separable finite extension is always simple. Also, a simple algebraic extension $\mathbf{F}(\alpha)$ is separable over $\mathbf{F}$ iff α is separable over $\mathbf{F}$. Whether "transitivity" holds for separable extensions, that is, if $\mathbf{F} \subset \mathbf{E} \subset \mathbf{K}$ are three fields such that $\mathbf{E}$ is a separable extension of $\mathbf{F}$, $\mathbf{K}$ is a separable extension of $\mathbf{E}$, is it true that $\mathbf{K}$ is a separable extension of $\mathbf{F}$? The answer is in the affirmative. This is, however, not true, if "being separable," is replaced by "being normal."

VIII.5 Definition. A field $\mathbf{F}$ is called perfect if each of its algebraic extensions is separable.

All fields of characteristic zero are perfect. A field $\mathbf{K}$ of characteristic $p \neq 0$ is perfect iff $\mathbf{K}^p = \mathbf{K}$, that is, iff every element of $\mathbf{K}$ is the pth root of some element of $\mathbf{K}$. In particular, finite fields are perfect.

IX. Finite Fields

Finite fields, that is, fields having a finite number of elements are of special interest in applications to problems in information theory and combinatorial geometry. A finite field, also called a Galois field, with q elements is generally denoted by $\mathbf{GF}(q)$. Coding theory makes a good deal of use of $\mathbf{GF}(2^n)$ by representing the elements of this field as strings of 0 and 1.

Let $\mathbf{F}$ be a finite field. Then the characteristic of $\mathbf{F}$ is prime p, and $\mathbf{F}$ contains a copy of $\mathbb{Z}/(p)$ (under the mapping sending $\bar{1}$ to 1). Regarding $\mathbf{F}$ as a vector space over $\mathbb{Z}/(p)$, it follows that $|\mathbf{F}| = p^n$ for some positive integer n. Furthermore, any nonzero element of a finite field $\mathbf{F}$ with p^n elements must satisfy $x^{p^n-1} = 1$, since $\mathbf{F} - \{0\}$ is a multiplicative group with $p^n - 1$ elements. Thus, $\mathbf{F}$ is a splitting field of $x^{p^n} - x$ over its prime subfield $\mathbf{F}_p$ ($\simeq \mathbb{Z}/(p)$). (A field is called prime if it has no proper subfields.) Thus, any two fields with p^n elements are isomorphic. Indeed, for each prime p, the roots of the polynomial $x^{p^n} - x$ over $\mathbb{Z}/(p)$ are all distinct and form a field with p^n elements. An important result of practical interest to information theorists is that there exists an irreducible polynomial of any given degree n over a finite field. This fact follows from a result that $\mathbf{F} - \{0\}$ is a multiplicative cyclic group and so if $\mathbf{E}$ is a finite field $\mathbf{F}$, then $\mathbf{E} = \mathbf{F}(\alpha)$, where α is the generator of $E - \{0\}$. Hence, if $m(x)$ is the minimal polynomial of α over $\mathbf{F}$, then $m(x)$ is an irreducible polynomial of degree $[\mathbf{E}: \mathbf{F}]$. Finite fields are perfect and each finite extension is a normal extension. The group of automorphisms of a finite field $\mathbf{GF}(p^n)$ is cyclic of degree n generated by ϕ where $\phi(x) = x^p$.

X. Fundamental Theorem of Galois Theory and Application

Let $\mathbf{E}$ be an extension of a field $\mathbf{F}$. $\mathbf{G}(\mathbf{E}/\mathbf{F})$ denotes the group of automorphisms of $\mathbf{E}$ leaving each element of $\mathbf{F}$ fixed.

X.1 Definition. Let $\mathbf{E}$ be a field and $\mathbf{H}$ a subgroup of the group of automorphisms of $\mathbf{E}$. Then the set $\mathbf{E}_\mathbf{H} = \{x \in \mathbf{E} | \sigma(x) = x, \text{ for all } \sigma \in \mathbf{H}\}$ is called the fixed field of the group of automorphisms $\mathbf{H}$.

Let $\mathbf{E}$ be a field, $\mathbf{G}$ a finite group of automorphisms of $\mathbf{E}$, and $\mathbf{F} = \mathbf{E}_\mathbf{G}$. The fundamental theorem of Galois theory states that

$$\mathbf{H} \to \mathbf{E}_\mathbf{H}, \qquad \mathbf{K} \to \mathbf{G}(\mathbf{E}/\mathbf{K})$$

define an order-preserving bijection between the set of subgroups of $\mathbf{G}$ and the set of subfields of $\mathbf{E}$ containing $\mathbf{F}$.

For a finite separable extension $\mathbf{E}$ of $\mathbf{F}$, the following three statements are equivalent: (i) $\mathbf{E}$ is a normal extension of $\mathbf{F}$, (ii) $\mathbf{F}$ is the fixed field of $\mathbf{G}(\mathbf{E}/\mathbf{F})$, and (iii) $[\mathbf{E}: \mathbf{F}] = |\mathbf{G}(\mathbf{E}/\mathbf{F})|$.

X.2 Definition. Let $f(x) \in \mathbf{F}[x]$ be a polynomial over $\mathbf{F}$, and $\mathbf{E}$ the splitting field of $f(x)$. The group $\mathbf{G}(\mathbf{E}/\mathbf{F})$ is called the Galois group of $f(x)$ over $\mathbf{F}$.

An important special case of the fundamental theorem of Galois theory gives that if $\mathbf{F}$ is a field of characteristic zero and if $\mathbf{G}(\mathbf{E}/\mathbf{F})$ is a Galois group of a polynomial $f(x) \in \mathbf{F}[x]$, then for any subfield $\mathbf{K}$ of $\mathbf{E}$ containing $\mathbf{F}$, the mapping $\mathbf{K} \to \mathbf{G}(\mathbf{E}/\mathbf{K})$ sets up a one-to-one correspondence from the set of subfields of $\mathbf{E}$ containing $\mathbf{F}$ to the subgroups of $\mathbf{G}(\mathbf{E}/\mathbf{F})$ such that (i) $\mathbf{K} = \mathbf{E}_{\mathbf{G}(\mathbf{E}/\mathbf{K})}$; (ii) for any subgroup $\mathbf{H}$ of $\mathbf{G}(\mathbf{E}/\mathbf{F})$, $\mathbf{H} = \mathbf{G}(\mathbf{E}/\mathbf{E}_\mathbf{H})$; (iii) $[\mathbf{E}: \mathbf{K}] = |\mathbf{G}(\mathbf{E}/\mathbf{K})|$, $[\mathbf{K}: \mathbf{F}]$ = index of $\mathbf{G}(\mathbf{E}/\mathbf{K})$ in $\mathbf{G}(\mathbf{E}/\mathbf{F})$; (iv) $\mathbf{K}$ is a normal extension of $\mathbf{F}$ iff $\mathbf{G}(\mathbf{E}/\mathbf{K})$ is a normal subgroup of $\mathbf{G}(\mathbf{E}/\mathbf{F})$; and (v) if $\mathbf{K}$ is a normal extension of $\mathbf{F}$, then $\mathbf{G}(\mathbf{K}/\mathbf{F}) \simeq \mathbf{G}(\mathbf{E}/\mathbf{F})/\mathbf{G}(\mathbf{E}/\mathbf{K})$.

The impetus to develop theory of fields in this direction originated from the necessity of finding necessary and sufficient conditions that a polynomial of degree 5 over $\mathbb{Q}$ is solvable by "radicals."

X.3 Definition. A polynomial $f(x) \in \mathbb{Q}[x]$ is said to be solvable by radical if its splitting field $\mathbf{E}$ is contained in $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$, where $\alpha_1^{n_1} \in \mathbb{Q}$, $\alpha_i^{n_i} \in \mathbb{Q}(\alpha_1, \ldots, \alpha_{i-1})$, $2 \leq i \leq r$, and n_i are positive integers. [An extension such as $\mathbb{Q}(\alpha_1, \ldots, \alpha_r)$ of $\mathbb{Q}$ is called a radical extension of $\mathbb{Q}$.]

In other words, a polynomial $f(x) \in \mathbf{F}[x]$ is solvable by radicals if each root of $f(x)$ can be obtained by using a finite sequence of operations of addition, subtraction, multiplication, division, and taking n_{ith} roots, starting with elements of $\mathbf{F}$.

The major application of the fundamental theorem of Galois theory is that if $\mathbf{F}$ is a field of

characteristic zero, then a polynomial $f(x) \in \mathbf{F}[x]$ is solvable by radicals iff its Galois group is solvable. [A group $\mathbf{G}$ is said to be solvable if there exists a finite chain

$$\{e\} = \mathbf{G}_0 \subset \mathbf{G}_1 \subset \cdots \subset \mathbf{G}_{n-1} \subset \mathbf{G}_n = \mathbf{G}$$

of subgroups of $\mathbf{G}$ such that $\mathbf{G}_i$ is a normal subgroup of $\mathbf{G}_{i+1}$ and $\mathbf{G}_{i+1}/\mathbf{G}_i$ is abelian, $0 \leq i \leq n-1$. It is well known that $\mathbf{S}_n$, the symmetric group of degree n, is not solvable if $n > 4$.) There were a number of attempts made earlier by many mathematicians including Gauss to find a formula for solving a polynomial of degree 5 by radicals. Abel in 1824 showed the impossibility for such a formula. It was in 1830 that Galois gave necessary and sufficient conditions in terms of solvability of the Galois group. For example, the Galois group of $x^4 - 2 \in \mathbb{Q}[x]$ is $\mathbf{D}_4$, the dihedral group of degree 4. Since $\mathbf{D}_4$ is a solvable group, $x^4 - 2$ is solvable, as it is well known that all polynomials of degree ≤ 4 are solvable by radicals. If one considers the polynomial $2x^5 - 5x^4 + 5 \in \mathbb{Q}[x]$ which has exactly three real roots in $[-1, 1]$, $[1, 2]$, $[2, 3]$, it follows that its Galois group is S_5. [If $f(x) \in \mathbb{Q}[x]$ is an irreducible polynomial over $\mathbb{Q}$ of degree p, p prime, having exactly two nonreal roots in $\mathbb{C}$, then the Galois group of $f(x)$ is $\mathbb{S}_p$.] Since S_5 is not a solvable group, it follows that $f(x)$ is not solvable by radicals.

In this connection a question naturally arises: given a finite group $\mathbf{G}$, does there exists $f(x) \in \mathbb{Q}[x]$ whose Galois group is $\mathbf{G}$. The general problem remains unsolved for about 100 years except for some special cases like symmetric groups, alternating groups, and solvable groups (thus groups of odd orders), for which the answer is known to be in the affirmative.

XI. Ruler and Compass Constructions

The theory of fields provides solution to many ancient geometric problems. Among such problems are

1. To construct, by ruler and compass, a square having the same area as that of a given circle.
2. To construct, by ruler and compass, a cube having twice the volume of a given cube.
3. To trisect a given angle by ruler and compass.
4. To construct, by ruler and compass, a regular polygon having n sides.

In the days of Euclid the only use of a ruler was to draw a line or line segment joining two given points and the only use of a compass was to draw a circle (or an arc) having a given point as its center and passing through another given point. Thus, a figure constructible by ruler and compass is completely determined by a set of points. If S is a nonempty subset of the Euclidean plane $\mathbb{R}^2$, then a line (or circle) is said to be constructible from S if it is the line through two distinct points in S (or it is the circle passing through a point of S with its center at another point of S). A point is constructible from S if it is a point common either to two distinct lines constructible from S, or to a line and a circle each constructible from S, or two distinct circles constructible from S. Let P_0 be a subset of $\mathbb{R}^2$ having at least two distinct points. Let P_1 be the set of all points constructible from P_0. Then $P_1 \supseteq P_0$. The process can be continued to obtain an infinite sequence (P_n), $n = 0, 1, 2, \ldots$ of subsets of $\mathbb{R}^2$ such that $P_{n-1} \supseteq P_n$, for each n. Let $P = \cup_n P_n$. A point, line, or circle constructible from P is usually called constructible from P_0.

A real number u is constructible from $\mathbb{Q}$ if the point $(u, 0)$ is constructible from $\mathbb{Q} \times \mathbb{Q} \subset \mathbb{R}^2$. Suppose that the coordinates of points of a certain subset P of $\mathbb{R}^2$ lie in a field $\mathbf{K}$. Then the equations of a line and a circle obtained from P are respectively of the form $ax + by + c = 0$, and $x^2 + y^2 + 2gx + 2fy + d = 0$, $a, b, c, f, g, d \in \mathbf{K}$. Thus, the coordinates of the point of intersection of two such distinct lines lie in K. Also, the coordinates of the points of intersection of such a line and such a circle or of two such distinct circles lie in $\mathbf{K}(\sqrt{\alpha})$, $\alpha > 0$, $\alpha \in \mathbf{K}$. It follows then that if $u \in \mathbb{R}$ is constructible from $\mathbb{Q}$, then there exists an ascending chain

$$\mathbb{Q} = \mathbf{K}_0 \subset \mathbf{K}_1 \subset \cdots \subset \mathbf{K}_n$$

of subfields of $\mathbb{R}$ such that (i) $u \in \mathbf{K}_n$, (ii) $\mathbf{K}_i = \mathbf{K}_{i-1}(\alpha_i)$, $\alpha_i^2 \in \mathbf{K}_{i-1}$, $1 \leq i \leq n$. Thus, $[\mathbf{K}_i : \mathbf{K}_{i-1}] \leq 2$, and so $[\mathbf{K}_n : \mathbb{Q}] = 2^m$. Therefore if $u \in \mathbb{R}$ is constructible from $\mathbb{Q}$ it is necessary that there exists a subfield $\mathbf{K}$ of $\mathbb{R}$ containing u such that $[\mathbf{K} : \mathbb{Q}]$ is a nonnegative power of 2. Another important fact is that the numbers constructible from $\mathbb{Q}$ form a subfield of $\mathbb{R}$.

Two immediate consequences of the above are (i) it is impossible to construct a square whose area is equal to the area of a circle of unit radius by using ruler and compass only, and (ii) it is impossible to construct a cube whose volume is two times the volume of a given cube of unit side by using ruler and compass only. For,

if a is the side of a square to be constructed in (i), $a^2 = \pi$. This implies a is not algebraic over $\mathbb{Q}$, since π is not algebraic over $\mathbb{Q}$ and so $[\mathbb{Q}(a): \mathbb{Q}] \neq 2^m$ for any nonnegative integer m. $\alpha \in \mathbb{C}$ is called a transcendental number if α is not algebraic over $\mathbb{Q}$. For (ii), if a is the length of the side of the cube to be constructed, then $a^3 = 2$, and so a satisfies the minimal polynomial $x^3 - 2$ over $\mathbb{Q}$. This means $[\mathbb{Q}(a): \mathbb{Q}] = 3 \neq 2^m$ for any nonnegative integer.

An angle is constructible if its vertex is a constructible point and if each of its sides contains a constructible point other than the vertex. It is an important consequence of the definition that an angle α is constructible iff $\cos \alpha$ is a constructible number or equivalently $(\cos \alpha, \sin \alpha)$, $(\cos \alpha, 0)$, or $(0, \sin \alpha)$ is a constructible point (of course iff $\sin \alpha$ is a constructible number). This yields $\alpha = 20°$ is not constructible, and so an angle of 60° cannot be trisected by ruler and compass (if $a = \cos 20°$, a satisfies the minimal polynomial $x^3 - 3x - 1 \in \mathbb{Q}[x]$).

Regarding the problem of constructing a regular polygon of n sides by ruler and compass, it is clear that if this can be done, then the angle $2\pi/n$ subtended at its center by a side must also be constructible. It is known by apealing to results in Galois theory that a regular polygon of n sides is constructible by ruler and compass iff $\phi(n)$ is a power of 2, where ϕ is the Euler function. In particular, 7-gon is not constructible by ruler and compass. If p is prime, then a p-gon is constructible by ruler and compass iff $p = 2^{2^h} + 1$ for some nonnegative integer h. (Primes of the form $2^m + 1$ are called Fermat primes.)

BIBLIOGRAPHY

Bhattacharya, P. B., Jain, S. K., and Nagpaul, S. R. (1986). "Basic Abstract Algebra." Cambridge University Press, London and New York.

Cohn, P. M. (1977). "Algebra," Vol. 2. Wiley, New York.

Cohn, P. M. (1982). "Algebra," Vol. 1, 2nd ed. Wiley, New York.

Jacobson, N. (1985). "Basic Algebra I." W. H. Freeman, New York.

Kaplansky, I. (1972). "Fields and Rings." Chicago University Press, Chicago.

McCoy, N. H. (1982). Rings and ideals. *In* "The Carus Mathematical Monographs," 6th ed. Math. Assoc. Amer. No. 8.