

§ 3. Commutative Rings, Applications to Groups

John R. Stallings, U. C. Berkeley

This particular chapter of notes is a short course in the theory of commutative rings. The group-theoretic applications are the following: (1) the matters related to “test sets” and the “Ehrenfeucht conjecture” about free monoids. This is an elementary application of the Hilbert Basis Theorem, that the polynomial ring over a Noetherian ring is Noetherian; it says that, in matters relating to homomorphisms of groups (or monoids) into monoids of $n \times n$ matrices (for fixed n), there are some finiteness results which involve the ascending chain property for equalizers. This subject grew out of “Formal Language Theory” and has had its major results since 1985. (2) Residual finiteness properties for groups (or monoids) of matrices over commutative rings. These properties were proved for matrices over fields by Malcev around 1940. The crucial fact here is that a field which is finitely generated as a ring is actually a finite field; I give an elementary proof of this which I received from Peter Shalen. Then, as Robert Coleman pointed out to me, the Krull Intersection Theorem will imply that every finitely generated comring is residually finite, and thus extend Malcev’s result to matrices over any comring, not necessarily a field.

I think that the modern (Goldman-Krull) version of Hilbert’s Nullstellensatz, which generalizes the result about finitely generated fields being finite, is a fascinating theorem; and so I include a proof of this here, as well as a few other tidbits about modules over p.i.d.s and factorial domains.

1. Ring definitions

• **1.1. The Category of Rings.** A *ring* (in general) consists of a set A ; two binary operations on A , that is, functions $+: A \times A \rightarrow A$ and $\cdot: A \times A \rightarrow A$; and additional things related to these operations, in particular a specific element $1 \in A$. $+$ and \cdot are supposed to be associative operations; and $+$ is commutative; and \cdot distributes over $+$ on both sides. $(A, +)$ is to be a commutative group; thus there is an “additive identity” denoted 0 , and to each $a \in A$, its “negative” $-a \in A$, such

that $x + 0 = x$ and $x + (-x) = 0$. (A, \cdot) is to be a monoid; in addition to the associative law, it satisfies the rule that $x \cdot 1 = 1 \cdot x = x$.

Rings are the objects in a Category; that is, there is a notion of a ring-map. Suppose that A and B are rings. A *ring-map* $\varphi: A \rightarrow B$ is a function on the underlying sets which preserves the ring structure; that is, $\varphi(x + y) = \varphi(x) + \varphi(y)$ and $\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y)$ and $\varphi(1) = 1$.

▷ EXERCISE 1.1: In the Category of rings, there is an initial object \mathbf{Z} , the ring of integers, and a terminal object 0 , the singleton or zero ring.

▷ EXERCISE 1.2: **Product of rings.** The product of a family of rings is easy to describe, whereas the coproduct can only be described with difficulty in any explicit case. If A and B are rings, then the product $A \times B$ consists of the cartesian product of the underlying sets of A and B , with the structure defined as follows:

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 \cdot a_2, b_1 \cdot b_2)$$

$$1_{A \times B} = (1_A, 1_B).$$

It might help to think of $A \times B$ as consisting of 2×2 -matrices

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \text{ for } a \in A, b \in B.$$

• **1.2. Ideals.** A *left ideal* I in a ring A is a subset of the underlying set of A , which forms a subgroup under $+$, and which satisfies the law

$$\forall a \in A, \forall x \in I, a \cdot x \in I.$$

We write this law simply: $A \cdot I \subset I$. Similarly, a *right ideal* J is an additive subgroup of A , such that $J \cdot A \subset J$; and a *two-sided ideal* K is a subset which is both a left ideal and a right ideal.

▷ EXERCISE 1.3: In the case of non-commutative rings, it is sometimes interesting to consider for a ring A , the *opposite ring* A^* . The ring A^* consists of the same underlying set as A , with the same $+$ and the same 1 , but with \cdot^* defined by: $a \cdot^* b = b \cdot a$. A left ideal in A is a right ideal in A^* , and vice versa.

2. Commutative ring definitions

Most of what these notes are about concerns commutative rings, that is, rings in which the operation \cdot is commutative: $x \cdot y = y \cdot x$. I am going to use the abbreviation “comring” for commutative ring. Comrings form a full subCategory of the Category of rings; the word “full” simply indicates that a comring-map is exactly the same as a ring-map whose source and target are commutative rings. The initial object \mathbf{Z} , the terminal object 0 , the product construction $A \times B$, are exactly the same in comrings as in rings.

** EXERCISE 2.1: **Tensor product of comrings.** The construction of coproduct of comrings, while not particularly simple, is easier to get your hands on than the coproduct of rings; it is a version of the notion of “tensor product.” Given two comrings A and B , we define an abelian group $A \otimes B$ to be the quotient group of the free abelian group on the ordered pairs $\{(a, b) \mid a \in A, b \in B\}$, by the subgroup generated by elements of the form $(a_1 + a_2, b) - (a_1, b) - (a_2, b)$ and $(a, b_1 + b_2) - (a, b_1) - (a, b_2)$. We define the binary operation of multiplication so that $(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2, b_1 b_2)$. The exercise is to verify that this makes sense, gives a commutative ring, and that the function (and its B analogue) $a \mapsto (a, 1)$ defines a ring-map from A into this ring, making this ring to have the property of coproduct of A and B in the Category of comrings.

* EXERCISE 2.2: The forgetful functor from comrings to sets has a left adjoint, which one would call the “free comring on a set X ”; this is, basically, what the notion of ring of polynomials concerns.

• **2.1. Domains.** A comring D in which $1 \neq 0$ and in which the set of non-zero elements is closed under multiplication is called an *integral domain* or simply a domain; that is, for all $x, y \in D$, if $x \neq 0$ and $y \neq 0$, then $xy \neq 0$.

• **2.2. Fields.** A comring F in which $1 \neq 0$ and the set of non-zero elements forms, multiplicatively, a group, is called a *field*; that is, for all $x \in F$, if $x \neq 0$, then there exists $y \in F$ such that $xy = 1$.

▷ EXERCISE 2.3: Every field is a domain. The ring of integers \mathbf{Z} is a domain which is not a field. Every finite domain is a field.

• **2.3. Factor ring.** Given an ideal I in a comring A , one can construct the *factor ring* (or *quotient ring*) A/I . The elements of A/I are the additive cosets of the

form $x + I$; addition and multiplication in A/I are defined in such a way that the natural map $A \rightarrow A/I$ is a ring-map. If $\varphi: A \rightarrow B$ is a ring-map, then the kernel of φ is

$$\ker \varphi = \{x \in A \mid \varphi(x) = 0\}.$$

This kernel is an ideal, and the image of φ , that is, $\varphi(A)$, is a subring of B . An elementary fact is that the quotient ring $A/\ker\varphi$ is isomorphic (ring-equivalent) to $\varphi(A)$. [If we do not assume that A is commutative, but only a ring, this paragraph is still true, provided that the word “ideal” is replaced by “two-sided ideal.”]

• **2.4. Prime.** An ideal P in a comring A is called a “prime ideal,” when A/P is an integral domain. This is the same as to say: $1 \notin P$; and $\forall x, y \in A \setminus P$, it is the case that $xy \in A \setminus P$.

• **2.5. Maximal.** An ideal M in a comring A is called a “maximal ideal,” when A/M is a field.

▷ EXERCISE 2.4: Equivalently: $1 \notin M$; and the only ideal of A which properly contains M is the entire comring A .

* EXERCISE 2.5: And so, clearly, every maximal ideal is a prime ideal. Furthermore, in a comring A , every ideal J which does not contain 1 is contained in at least one maximal ideal (this uses Zorn’s Lemma).

• **2.6. Generating set.** An ideal I in a comring A is said to be *generated* by a subset X of A , if the smallest ideal containing X is I . We call I a *finitely generated* ideal if it is generated by some finite set. We call I the *principal ideal* generated by x , if I is generated by the singleton set $\{x\}$. We denote the ideal of A generated by the subset X by $\langle X \rangle$, or by AX , or by $\langle X \rangle_A$.

▷ EXERCISE 2.6: Every element of I can be written as a finite sum $\sum a_i x_i$, where $a_i \in A$ and $x_i \in X$; the set of these finite sums forms the ideal I .

• **2.7. Noetherian.** A comring is said to be *Noetherian*, when every ideal in it is finitely generated.

* EXERCISE 2.7: **Noetherian rings.** This is equivalent to the *ascending chain condition* on ideals: If $I_1 \subset I_2 \subset I_3 \subset \dots$ is a sequence of ideals, each containing the previous one, and if J is the union of these ideals, then there is some n such that $I_n = J$. (Warning: It is a little bit tricky to show that the ascending chain condition on ideals implies that every ideal is finitely generated.)

• **2.8. PID.** A *principal ideal domain* or p.i.d., is an integral domain in which each ideal is principal. Clearly, every p.i.d. is Noetherian. The standard example of a p.i.d. is \mathbf{Z} , in which the Euclidean algorithm gives a way to show that each ideal is principal. This is generalized by the notion of a *Euclidean domain*, which is an integral domain D with a “Euclidean valuation” $e: D \rightarrow \Omega$ on it. Here, Ω is a well-ordered set; let us call the least element of Ω by the name “0.” The rules for a Euclidean valuation are (1) $e(x) = 0 \iff x = 0$; (2) $\forall y \neq 0, e(xy) \geq e(x)$; (3) $\forall x, y \in A$, if $y \neq 0$, then $\exists q, r \in A \ni x = yq + r$ and $e(r) < e(y)$.

* EXERCISE 2.8: Suppose that we have a Euclidean domain, defined as above. Show that every ideal is principal. (There are p.i.d.s which have no Euclidean structure. An example is the factor ring of the polynomial ring in one indeterminate, $\mathbf{Z}[x]$, by the ideal generated by $x^2 - x + 5$. It can be fun to figure out why this is so.)

• **2.9. Polynomials.** Given a comring A , we can describe the *polynomial ring* $A[x]$. This is the set of formal expressions of finite length

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

with $a_i \in A$; we add and multiply these expressions in the customary way. We can symbolically imagine $A[x]$ to consist of

$$A + Ax + Ax^2 + \cdots + Ax^n + \cdots.$$

The *degree* of $f(x)$ above is said to be $\leq n$, and to equal n when its *leading coefficient* a_n is unequal to 0.

▷ EXERCISE 2.9: If D is an integral domain, then $D[x]$ is an integral domain. If F is a field, then $F[x]$ is a Euclidean domain. Every factor ring A/I of a Noetherian ring A by an ideal is Noetherian.

It is *not* true that every subring of a Noetherian ring is Noetherian. Here is one of the easiest examples. As we shall show in the section on the Hilbert Basis Theorem, the ring $\mathbf{Z}[x]$ is Noetherian. It contains the subring B consisting of all polynomials

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

such that the coefficients a_i , for $i > 0$, are divisible by 2; we can describe this pictorially:

$$B = \mathbf{Z} + 2\mathbf{Z}x + 2\mathbf{Z}x^2 + \cdots + 2\mathbf{Z}x^n + \cdots.$$

In B , the ideal of those polynomials $f(x)$, having constant term $a_0 = 0$, is generated as an ideal in B by the infinite set

$$\{2x, 2x^2, 2x^3, \dots, 2x^n, \dots\}$$

and by no proper subset thereof. And so this ideal in B is not finitely generated. A peculiar phenomenon occurs here: This same set is an ideal in $\mathbf{Z}[x]$, where it is the principal ideal generated in $\mathbf{Z}[x]$ by $2x$.

3. Modules

If A is a ring and V is an abelian group (additive notation), then we call V a *left A -module*, when we have defined a function $\mu: A \times V \rightarrow V$, denoted by the slightly confusing notation $ax = \mu(a, x)$, satisfying the rules: $1x = x$, $(ab)x = a(bx)$, $(a + b)x = (ax) + (bx)$, and $a(x + y) = (ax) + (ay)$.

There is also the notion of *right A -module*, which is related to a ring-map from A to the opposite ring of $\text{End}(G)$.

A *map of A -modules* $\varphi: V \rightarrow W$ is an abelian-group-map such that, for all $a \in A$ and all $x \in V$, it is the case that $\varphi(ax) = a\varphi(x)$. Thus we get a Category of A -modules.

* EXERCISE 3.1: To each abelian group G , we can associate a ring $\text{End}(G)$ (the “endomorphisms” of G), which consists of all the abelian-group-maps $\alpha: G \rightarrow G$, where we define $(\alpha\beta)(x) = \alpha(\beta(x))$ and $(\alpha + \beta)(x) = \alpha(x) + \beta(x)$. The definition of $+$ in $\text{End}(G)$ is where the fact that G is abelian is used. Then a structure of left A -module on G is exactly the same as a ring-map $A \rightarrow \text{End}(G)$.

▷ EXERCISE 3.2: **Category of modules.** The category-theoretic constructions of product and coproduct correspond to the notions which we customarily call “direct product” and “direct sum.” In particular, the product of V and W is denoted $V \times W$, and consists of the cartesian product of the underlying sets with operations defined like this: $(x, y) + (u, v) = (x + u, y + v)$ and $a(x, y) = (ax, ay)$. Similarly, we define V^n by $V^1 = V$ and $V^{n+1} = V^n \times V$. The coproduct and product have a canonical equivalence between them when there is a finite number of factors. The ring A itself is a left A -module.

▷ EXERCISE 3.3: The notion in general algebraic Categories of the subthing generated by a subset is that this is the smallest subthing containing the given subset.

There is generally another, internal, way to describe this. In the Category of A -modules, the submodule of V generated by a subset $X \subset V$ consists of those elements of V which can be written as a linear combination of the elements of X , with coefficients in A ; i.e., as some finite sum $\sum a_i x_i$, for $a_i \in A$ and $x_i \in X$. Thus, a sub- A -module of A itself is the same as a left ideal in A , and a set generating this ideal as an ideal in A is the same as a set generating it as an A -module. Similarly, we can speak of finitely generated modules. A module generated by a single element is called *cyclic*.

If V is a left A -module and S is a submodule of V , then there is a quotient module V/S . A cyclic module is A -module-equivalent to a module of the form A/I , where I is a left ideal of A .

• **3.1. Free modules.** The forgetful functor from left A -modules to sets has a left adjoint; given a set X , this produces $\Phi(X)$, the *free A -module on the set X* ; this consists of the coproduct of copies of A , one for each element of X . The set X is (if $1 \neq 0$ in A) embedded in the underlying set of $\Phi(X)$, and it is called a *free basis* of $\Phi(X)$. If we have, for example, an A -module equivalence $\varphi: A^k \rightarrow A^n$, then the image of a free basis of A^k is also a free basis of A^n .

* EXERCISE 3.4: It is *not* in general true that under these circumstances $k = n$. Find an example of a (non-commutative) ring A such that the left modules A and $A \oplus A$ are isomorphic. For example, you can do this with the ring A of $\infty \times \infty$ integer matrices which are row-finite (each row having only a finite number of non-zero entries).

• **3.2. Invariant basis number.** But suppose that, whenever, for finite k and n , A^k and A^n are isomorphic A -modules, then it follows that $n = k$. In this case, we say that the ring A has the *invariant basis number* property. It turns out that every commutative ring in which $1 \neq 0$ has the invariant basis number property; this can be proved using determinants or exterior algebras. The preferred method is to use the exterior algebra method.

• **3.3. Matrices.** Given an A -module-map $\varphi: A^k \rightarrow A^n$, we can describe a $k \times n$ matrix of elements of A as follows: First, we represent the elements of A^k as k -tuples (a_1, a_2, \dots, a_k) ; and similarly for A^n ; this amounts to choosing free bases of these two modules. We define ϵ_i to be the array consisting of 0 except in the i th place, where it is 1; this ought to have another symbol k or n attached to it, so that

one can determine whether it belongs to A^k or to A^n , but we hope not to become confused. Now what we have is that $\{\epsilon_1, \dots, \epsilon_k\}$ is a free basis of A^k , and there is a similar basis of A^n . Now, φ is exactly determined by the values $\varphi(\epsilon_i)$:

$$\varphi(\epsilon_i) = \sum_{j=1}^n a_{ij} \epsilon_j.$$

The matrix (a_{ij}) is the matrix which we say is associated to φ . We can multiply matrices in a simple generalization of the way in which we are accustomed to multiply matrices of real numbers; the product of matrices is associated to the composition of the associated A -module-maps.

• **3.4. How ring facts imply module facts.** There is another construction dealing with modules over a comring, which often allows one to deduce facts about modules from theorems about comrings. Let V be a module over the comring A ; we define a new comring denoted by $A+V$ thus: The underlying set of $A+V$ is the cartesian product $A \times V$; the operations of addition and multiplication are defined thus:

$$(a, x) + (b, y) = (a + b, x + y)$$

$$(a, x) \cdot (b, y) = (ab, ay + bx)$$

▷ **EXERCISE 3.5:** What this achieves is to make V , as the subset $0 \times V \subset A + V$, into an ideal in the ring $A + V$; as an ideal there it is generated by any subset which generates it as an A -module. A submodule of V becomes the same as an ideal of $A + V$ which is a subset of $0 \times V$. Each prime ideal of $A + V$ contains V and intersects $A \times 0 = A$ in a prime ideal. This correspondence between prime ideals of $A + V$ and of A is bijective.

Here are some interesting things about prime ideals, written in the form of a couple of exercises:

* **EXERCISE 3.6: The prime spectrum.** Let A be a comring. The set of all prime ideals in A will be called the “prime spectrum” of A , or $\text{Spec}(A)$. Let I be any ideal in A ; define $C(I)$ to be the set of all prime ideals which contain I . Note that

$$C(A) = \emptyset \quad \text{and} \quad C(\{0\}) = \text{Spec}(A).$$

If I and J are two ideals, and P is a prime ideal, then if $I \cdot J \subset P$, it follows that either $I \subset P$ or $J \subset P$. Therefore:

$$C(I) \cup C(J) \subset C(I \cap J) \subset C(I \cdot J) \subset C(I) \cup C(J).$$

If $\{I_\alpha\}$ is any set of ideals, and J is the ideal generated by their union, then:

$$\bigcap C(I_\alpha) = C(J).$$

Therefore, the sets $\{C(I)\}$ satisfy the axioms for closed sets which make $\text{Spec}(A)$ into a topological space. What does it mean for a point P of $\text{Spec}(A)$ to form a singleton closed set? — Show that if you have any family of closed sets, such that the intersection of any finite number of this family is non-empty, then the intersection of the entire family contains such a closed point. Thus, $\text{Spec}(A)$ is a compact topological space, generally non-Hausdorff. A ring-map $f: A \rightarrow B$ yields a continuous function $f^*: \text{Spec}(B) \rightarrow \text{Spec}(A)$. Thus, Spec is a functor from the dual of the Category of comrings to the Category of compact topological spaces.

* **EXERCISE 3.7: Boolean rings and their spectra.** Let A be a ring in which $\forall a \in A, a \cdot a = a$. Such a ring is called a Boolean ring. Show that, in a Boolean ring it is always true that $a \cdot b = b \cdot a$, so that every Boolean ring is in fact a comring. Every prime ideal in a Boolean ring is maximal, and the factor ring is isomorphic to the field consisting of only two elements. Let $C(I)$ be a closed set of $\text{Spec}(A)$ and let M be an element of $\text{Spec}(A)$ which is not in $C(I)$; then there exists $a \in I$ with $a \notin M$. In which case, $C(I) \subset C(\langle a \rangle)$ and $M \in C(\langle 1+a \rangle)$ and $C(\langle a \rangle) \cup C(\langle 1+a \rangle) = \text{Spec}(A)$ and $C(\langle a \rangle) \cap C(\langle 1+a \rangle) = \emptyset$. In other words, the prime spectrum of a Boolean ring is a totally disconnected, compact Hausdorff space; this is sometimes called the Stone space of the Boolean ring. Show how the elements of the original ring correspond exactly to those subsets of the Stone space which are both open and closed.

4. The Hilbert basis theorem

4.1. Hilbert Basis Theorem. *If A is a Noetherian ring, then $A[x]$ is Noetherian.*

Proof: Let I be an ideal in $A[x]$. For $k = 0, 1, \dots$, let I_k be the set of those elements of A which occur as leading coefficients of elements of I of degree k , together with the element 0. Then

$$I_0 \subset I_1 \subset \dots$$

is an ascending sequence of ideals in A . Since A is Noetherian, this sequence stabilizes at some point:

$$I_n = \bigcup_{i=0}^{\infty} I_i$$

Now, for $i = 0, \dots, n$, let $\{a_{i1}, \dots, a_{im_i}\}$ be a finite set of generators of I_i as an A -ideal. Let $f_{ij}(x) \in A[x]$ be a polynomial of degree i , in which the coefficient of x^i is a_{ij} .

The claim then is that

$$X = \{f_{01}, \dots, f_{0m_0}, f_{11}, \dots, f_{1m_1}, \dots, f_{n1}, \dots, f_{nm_n}\}$$

is a set of generators of I as $A[x]$ -ideal. Let $g(x)$ be an arbitrary element of I , with leading term bx^k , where $b \in A$ and the degree of g is k . If $k \leq n$, then a linear combination (coefficients in A) of $\{f_{k1}, \dots, f_{km_k}\}$ can be subtracted from g to reduce its degree. If $k > n$, then a linear combination (coefficients in $A \times x^{k-n}$) of $\{f_{n1}, \dots, f_{nm_n}\}$ can be subtracted to reduce the degree. Continuing thus, reducing the degree of g a finite number of times, we express g as a linear combination (coefficients in $A[x]$) of the elements of X . ■

It follows that the rings $\mathbf{Z}[x_1, \dots, x_n]$ and $F[x_1, \dots, x_n]$ for F any field, are Noetherian; and all quotient rings of these are Noetherian. A special example: Every finitely generated ring (quotient of $\mathbf{Z}[x_1, \dots, x_n]$) is Noetherian.

Note that it is possible to invent ideals in these examples without being able, *algorithmically*, to find a set of generators. Even in \mathbf{Z} , the ideal I generated by

$$X = \{n \geq 3 \mid \exists x, y, z \in \mathbf{Z}^+ \ni x^n + y^n = z^n\}$$

is generated by one element, but until recently, we didn't know if that element is 0 or 1 or 6,421,299.

And, for instance, if P , Q , and R denote the following 2×2 matrices in $\mathbf{Z}[x_1, \dots, x_{12}] = A$:

$$P = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \quad Q = \begin{pmatrix} x_5 & x_6 \\ x_7 & x_8 \end{pmatrix} \quad R = \begin{pmatrix} x_9 & x_{10} \\ x_{11} & x_{12} \end{pmatrix}$$

and if I is the ideal generated by the entries in the matrices

$$PQR^n - R^nQP,$$

we could spend years trying to find a finite set of generators of this ideal.

Another thing to note is that we can't give an *a priori* bound on the number of generators of an ideal. In $\mathbf{Z}[x]$, the ideal generated by

$$\{2^n, 2^{n-1}x, \dots, 2^{n-k}x^k, \dots, x^n\}$$

cannot be generated by fewer elements.

The construction, for a module V , of the ring $A + V$, yields this fact:

4.2. Theorem. *If A is Noetherian and V is a finitely generated A -module, then $A + V$ is Noetherian and every submodule of V is finitely generated. ■*

5. Examples of matrix monoids

A number of monoids and groups can be embedded in monoids of matrices over commutative rings. These examples therefore can be studied with the help of the theory of commutative rings. We first list some examples.

• **5.1. Free Monoids.** Let A be a set. We think of A as an *alphabet* and construct the set of *words* in A . A word is a finite ordered sequence of elements of A , that is, an element of some finite Cartesian product $A \times \cdots \times A$; the number of factors of this product, that is, the number of letters in the word w , is called the *length* of w . There is a unique word of length 0, the *empty word*, denoted 1. If u and v are words, of lengths k and n , we can concatenate them to produce a word $u \cdot v$ of length $k + n$. This makes the set of words in A , with the binary operation of concatenation, into a monoid A^* called the free monoid on the set A .

▷ **EXERCISE 5.1: Countable free monoids are contained in free monoid on two elements.** The submonoid of $\{a, b\}^*$ generated by the elements $\{ab^n : n \geq 1\}$ is monoid-equivalent to the free monoid on a countable set.

▷ **EXERCISE 5.2: Embedding a free monoid in 2×2 upper triangular matrices.** Now, let $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. To the element $w \in A^*$, we can associate its length n and the integer d whose decimal notation is w , and we have $0 \leq d < 10^n$; conversely, given n and d satisfying this inequality, there is a unique word $w \in A^*$ associated to this pair. Now, to w we associate the matrix

$$\varphi(w) = \begin{pmatrix} 1 & d \\ 0 & 10^n \end{pmatrix} \in \mathbf{M}_2(\mathbf{Z})$$

and compute that if to u are associated n_u and d_u , and to v are associated n_v and d_v , then

$$\varphi(u) \cdot \varphi(v) = \begin{pmatrix} 1 & d_u \\ 0 & 10^{n_u} \end{pmatrix} \cdot \begin{pmatrix} 1 & d_v \\ 0 & 10^{n_v} \end{pmatrix} = \begin{pmatrix} 1 & 10^{n_v}d_u + d_v \\ 0 & 10^{n_u+n_v} \end{pmatrix} = \varphi(u \cdot v).$$

Thus, every countable free monoid can be embedded in $\mathbf{M}_2(\mathbf{Z})$.

• **5.2. Free groups.** Given a set X , we invent a disjoint set \overline{X} in one-to-one correspondence with X , the element $\overline{a} \in \overline{X}$ corresponding to $a \in X$. We define $\overline{\overline{a}} = a$. In $\{X \cup \overline{X}\}^*$, let \sim be the smallest equivalence relation such that

$$\forall a \in X \cup \overline{X}, a \cdot \overline{a} \sim 1, \text{ and}$$

$$\forall p, q, u, v \in \{X \cup \overline{X}\}^*, \text{ if } p \sim q, \text{ then } u \cdot p \cdot v \sim u \cdot q \cdot v.$$

On the resulting set of \sim -equivalence-classes, there is an inherited binary operation which is associative. This is a group, denoted $F(X)$, the free group on the set X . The inverse of the \sim -class containing $w = a_1 \cdot a_2 \cdots a_n$ is the \sim -class containing $\overline{w} = \overline{a_n} \cdots \overline{a_2} \cdot \overline{a_1}$.

A *reduced word* in $\{X \cup \overline{X}\}^*$, is an element $w = a_1 \cdot a_2 \cdots a_n$, such that, for all $i \in [1, n-1]$, we have $a_{i+1} \neq \overline{a_i}$. In order to embed $F(X)$ in $\mathbf{M}_2(\mathbf{Z})$, it is enough to define a monoid-map $\alpha: \{X \cup \overline{X}\}^* \rightarrow \mathbf{M}_2(\mathbf{Z})$, such that $\alpha(\overline{a}) = (\alpha(a))^{-1}$, and such that if w is a reduced word of length greater than 0, then $\alpha(w)$ is not the identity matrix. This will, incidentally, imply that each \sim -class contains a unique reduced word. To do this, for $X = \{a, b\}$, and hence, by an argument not dissimilar to that for free monoids, for countable free groups, we shall use a trick due to Macbeath:

Divide the upper half of the xy -plane into four 45° sectors, counterclockwise from the positive x -axis, $B, A, \overline{B}, \overline{A}$, and extend these symmetrically through the origin. Draw this yourself.

Let

$$\alpha(a) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \alpha(b) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

We can compute that

$$\alpha(a)(A \cup B \cup \overline{B}) \subset A$$

$$\alpha(b)(A \cup \overline{A} \cup B) \subset B$$

$$\alpha(a^{-1})(\overline{A} \cup B \cup \overline{B}) \subset \overline{A}$$

$$\alpha(b^{-1})(A \cup \overline{A} \cup \overline{B}) \subset \overline{B}.$$

In other words, α takes each $x \in \{a, b, a^{-1}, b^{-1}\}$ into a linear map of the plane which takes all the plane except \overline{X} into X . It then follows that any reduced word $w = x_1 \cdots x_n$ of length $n \geq 1$ goes by α into a linear map which takes 3/4 of the plane, all except $\overline{X_n}$, into 1/4 of the plane, namely X_1 , and so α is not the identity.

▷ EXERCISE 5.3: You can do something similar with the matrices

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

But not with

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

* EXERCISE 5.4: You can also show that, for almost every real number λ (in fact, for all but a countable number of λ), the matrices

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$$

generate a free group of rank 2.

• **5.3. Other groups of interest.** There are a lot of hyperbolic 2-manifolds and hyperbolic 3-manifolds. This means that their fundamental groups embed, respectively, in $PSL_2(\mathbf{R})$ and in $PGL_2(\mathbf{C})$. In this notation, “ GL ” means “invertible matrices,” and “ SL ” means “matrices with determinant 1,” and “ P ” means the quotient by the center (= the constant diagonal matrices).

Now, for instance, $GL_2(\mathbf{C})$ acts by conjugation on the entire set of 2×2 complex matrices, which themselves form a vector space of dimension 4 over \mathbf{C} . This defines a group-map from $GL_2(\mathbf{C})$ into $GL_4(\mathbf{C})$, whose kernel is exactly the center of the source group. It thus embeds $PGL_2(\mathbf{C})$ in $GL_4(\mathbf{C})$.

6. Test sets

• **6.1. Equalizer in the Category of monoids.** Let $\alpha, \beta: S \rightarrow A$ be a pair of monoid-maps. There is a concrete description of the *equalizer* of α and β , as follows:

$$\text{Eq}(\alpha, \beta) = \{x \in S : \alpha(x) = \beta(x)\}.$$

• **6.2. Test sets.** Let \mathcal{A} be a class of monoids, and let S be a monoid, not necessarily in \mathcal{A} . Suppose that P is a subset of S . We call a subset $Q \subset P$ a *test set* (for P in S , relative to \mathcal{A}), when, for every A in \mathcal{A} and for every pair of monoid-maps $\alpha, \beta: S \rightarrow A$, if $Q \subset \text{Eq}(\alpha, \beta)$, then $P \subset \text{Eq}(\alpha, \beta)$. In other words, to test if two monoid-maps to a monoid in \mathcal{A} agree on P , it suffices to check that they agree on Q .

We shall call \mathcal{A} an *Ehrenfeucht class* of monoids when it is the case that for every finitely generated *free* monoid W and every subset $P \subset W$, there is a finite test set $Q \subset P$, for P in W , relative to \mathcal{A} .

* EXERCISE 6.1: If \mathcal{A} is an Ehrenfeucht class, then for every finitely generated (not necessarily free) monoid S and every subset $P \subset S$, there is a finite test set $Q \subset P$, for P in S , relative to \mathcal{A} .

Map a finitely generated free monoid W onto S , find a finite test set Q for the inverse image of P , and then the image of Q is a test set for P .

* EXERCISE 6.2: Given a monoid S , define $\mathbf{E}_{\mathcal{A}}(S)$ to be the set of equalizers of monoid-maps from S to elements of \mathcal{A} :

$$\mathbf{E}_{\mathcal{A}}(S) = \{ T \subset S : \exists A \in \mathcal{A}, \exists \alpha, \beta: S \rightarrow A \ni T = \text{Eq}(\alpha, \beta) \}.$$

If S is a finitely generated monoid and \mathcal{A} is an Ehrenfeucht class, then $\mathbf{E}_{\mathcal{A}}(S)$ satisfies the ascending chain condition. That is, if

$$T_1 \subset T_2 \subset \cdots \subset T_n \subset \cdots,$$

where $T_i \in \mathbf{E}_{\mathcal{A}}(S)$, then the union

$$T_{\infty} = \bigcup_1^{\infty} T_i$$

is equal to some T_n . The point is, that there exists a finite test set for T_{∞} , which is then contained in some T_n , which therefore contains T_{∞} .

* EXERCISE 6.3: Let S be a monoid, and $\alpha: S \rightarrow S$ a self-monoid-map. The *fixed set* and the *recurrent set* are defined as follows:

$$\text{Fix}(\alpha) = \{ x \in S : x = \alpha(x) \}$$

$$\text{Rec}(\alpha) = \{ x \in S : \exists n \geq 1 \ni x = \alpha^n(x) \}.$$

Now, if S is a submonoid of A , then $\text{Fix}(\alpha)$ can be described as an equalizer of a pair of maps from S to A . Furthermore, $\text{Rec}(\alpha)$ is the union of an ascending chain:

$$\text{Fix}(\alpha) \subset \text{Fix}(\alpha^2) \subset \cdots \subset \text{Fix}(\alpha^{n!}) \subset \cdots.$$

Conclude that if S is any finitely generated submonoid of a monoid A belonging to some Ehrenfeucht class, and if $\alpha: S \rightarrow S$ is any self-map of S , then there exists a positive integer n , such that $\text{Rec}(\alpha) = \text{Fix}(\alpha^n)$.

• **6.3. Matrix monoids are Ehrenfeucht.** Now, let \mathcal{A}_n be the class of monoids of the form $\mathbf{M}_n(A)$ for A an arbitrary comring.

6.4. Theorem. \mathcal{A}_n is an Ehrenfeucht class of monoids.

Proof: Consider a free monoid of finite rank, $W = \{x_1, \dots, x_r\}^*$. We now look at the polynomial ring $R = \mathbf{Z}[y_{ij,k}, z_{ij,k}]$, where $1 \leq i, j \leq n$ and $1 \leq k \leq r$. This is, by the Hilbert Basis Theorem, a Noetherian ring.

We define monoid-maps $\varepsilon_1, \varepsilon_2: W \rightarrow \mathbf{M}_n(R)$ by

$$\varepsilon_1(x_k) = [y_{ij,k}] \quad \text{and} \quad \varepsilon_2(x_k) = [z_{ij,k}].$$

Here by $[y_{ij,k}]$ is meant the matrix whose ij -th entry is the symbol $y_{ij,k}$.

Now, of course, $\mathbf{M}_n(R)$ also has an additive structure, with respect to which one can determine, for any $w \in W$, a matrix $\varepsilon_1(w) - \varepsilon_2(w)$.

Let, for any subset $P \subset W$, the ideal I_P of R be the ideal generated by the set of entries of the matrices in the set

$$D_P = \{ \varepsilon_1(w) - \varepsilon_2(w) : w \in P \}.$$

Since R is a Noetherian ring, I_P is generated as an ideal by a finite set. Therefore, there is a finite subset $Q \subset P$, such that I_P is generated by the set of entries of the matrices in the set

$$D_Q = \{ \varepsilon_1(w) - \varepsilon_2(w) : w \in Q \}.$$

The claim is that Q is the desired finite test set for P .

Let $\alpha, \beta: W \rightarrow \mathbf{M}_n(A)$ be a pair of monoid-maps. These determine a comring-map $\varphi_{\alpha, \beta}: R \rightarrow A$, by the formulas

$$\varphi_{\alpha, \beta}(y_{ij,k}) = \text{the } ij\text{-th entry in the matrix } \alpha(x_k).$$

$$\varphi_{\alpha, \beta}(z_{ij,k}) = \text{the } ij\text{-th entry in the matrix } \beta(x_k).$$

Now, let Y_k be the matrix $[y_{ij,k}]$ in $\mathbf{M}_n(R)$, and let Z_k be the matrix $[z_{ij,k}]$. Thus, $\varepsilon_1(x_k) = Y_k$ and $\varepsilon_2(x_k) = Z_k$. And $\mathbf{M}_n(\varphi_{\alpha, \beta})(Y_k) = \alpha(x_k)$, and so on. Thus, on the generating set $\{x_k\}$, and hence on all of W , we have

$$\mathbf{M}_n(\varphi_{\alpha, \beta})\varepsilon_1 = \alpha \quad \text{and} \quad \mathbf{M}_n(\varphi_{\alpha, \beta})\varepsilon_2 = \beta.$$

Thus, if we assume that α agrees with β on Q , then $\varphi_{\alpha, \beta}(\delta) = 0$ on all entries δ of all the matrices in the set D_Q . Since $\varphi_{\alpha, \beta}$ is a ring-map, we conclude that $\varphi_{\alpha, \beta}$ is zero on the ideal in R generated by the entries of all these matrices. That is,

$$\varphi_{\alpha, \beta}(I_P) = \{0\}.$$

But this implies, working back through all this, that α and β agree on P . ■

* EXERCISE 6.4: For a little more along these lines, look at Stallings, “Finiteness properties of matrix representations,” *Ann. of Math.* **124** (1986). Or, better, come up with your own generalizations of this matter. There is also an interesting paper by Simon Thomas *Proc. AMS* (10 or 15 years ago) which uses this idea to prove a conjecture of P. Scott’s.

* EXERCISE 6.5: Prove the “Ehrenfeucht Conjecture.” This is that if \mathcal{A} is the class of free monoids, then for every subset P of a finitely generated free monoid W , there is a finite test set Q for P in W , relative to \mathcal{A} . The point is, we only need to consider countable free monoid targets, and they are all embeddable in $\mathbf{M}_2(\mathbf{Z})$, which is contained in the Ehrenfeucht class \mathcal{A}_2 . You may also want to look up the note by Albert and Lawrence in *Theoretical Computer Science* (1986), where the Ehrenfeucht Conjecture was first proved, and other articles in CS-world with the word “Ehrenfeucht” in the title.

7. Determinants

The definition of determinant of an $n \times n$ matrix M over a comring A is a bit of a mess. The neat way seems to be to define the exterior algebra of a module and go on from there. The short but incomprehensible way is to define it by a formula.

We outline the exterior algebra theory in some exercises:

* EXERCISE 7.1: **Exterior algebra of a module.** Let M be a module over a comring A . We define $E_0(M) = A$. For $n \geq 1$, we first consider the set M^n consisting of n -tuples of elements of M . We consider the free A -module on this set; call it $F(M^n)$. Look at the submodule S_n generated by all the following sorts of expressions:

For some $i \in [1, n]$; for $j \neq i$, suppose $m_j = m'_j$; and $m'_i = am_i$:

$$a \cdot (m_1, \dots, m_n) - (m'_1, \dots, m'_n).$$

For some $i \in [1, n]$; for $j \neq i$, suppose $m_j = m'_j = m''_j$; and $m_i = m'_i + m''_i$:

$$(m_1, \dots, m_n) - (m'_1, \dots, m'_n) - (m''_1, \dots, m''_n).$$

For some $i, j \in [1, n]$, suppose $i \neq j$ and $m_i = m_j$:

$$(m_1, \dots, m_n).$$

We then define $E_n(M) = F(M^n)/S_n$. The element of $E_n(M)$ represented by (m_1, \dots, m_n) is conventionally denoted by $m_1 \wedge \dots \wedge m_n$. This is defined so that \wedge satisfies tensor-product types of relations plus the additional kind that says, when a factor is duplicated, the result is 0.

* **EXERCISE 7.2: The structure, in some cases, of the exterior algebra.** We note that the sign changes when two factors in $m_1 \wedge \cdots \wedge m_n$ are interchanged. For instance,

$$m_1 \wedge m_2 + m_2 \wedge m_1 = (m_1 + m_2) \wedge (m_1 + m_2) - m_1 \wedge m_1 - m_2 \wedge m_2 = 0.$$

\wedge induces an associative, distributive multiplication

$$E_k(M) \wedge E_n(M) \rightarrow E_{k+n}(M),$$

which has the additional property that, for $\alpha \in E_k(M)$, $\beta \in E_n(M)$, we have $\alpha \wedge \beta = (-1)^{kn} \beta \wedge \alpha$. We therefore call $E_*(M)$ an *anti-commutative graded ring*.

If M is a free A -module with basis $\{x_1, x_2, \dots\}$, then $E_k(M)$ is a free A -module with basis

$$\{x_{i_1} \wedge x_{i_2} \wedge \cdots \wedge x_{i_k} : i_1 < i_2 < \cdots < i_k\}.$$

* **EXERCISE 7.3: Proof of Invariant basis number for comrings.** Hence, if M is free with a basis of exactly n elements, it follows that $E_n(M)$ is isomorphic, as an A -module, to A , and that $E_{n+1}(M) = 0$. [This shows that every comring in which $1 \neq 0$ has the invariant basis number property.]

* **EXERCISE 7.4: Definition of determinant.** Every endomorphism $\phi: M \rightarrow M$ yields a map of graded rings $\varphi_*: E_*(M) \rightarrow E_*(M)$. In the case that M is free on a basis of n elements, then φ_n maps $E_n(M) \approx A$ to itself and is therefore multiplication by an element of A called the *determinant* of φ : $\varphi_n(\alpha) = d \cdot \alpha$ and $d = \det(\varphi)$.

* **EXERCISE 7.5:** There is a proof using determinants (supposing you know something about determinants without defining the exterior algebra first), that every non-trivial comring has the invariant basis number property.

In standard books, you can find more about the determinant. In particular, there is a formula for the determinant of an $n \times n$ matrix.

8. Properties of determinants

Here is a list of some of the properties of determinants. In the following, M , P , Q , etc., denote square matrices with coefficients in the comring A .

1. $\det(M) \in A$.
2. $\det(PQ) = \det(P) \cdot \det(Q)$.
3. Let $I = (\delta_{ij})$, where $\delta_{ij} = 0$ if $i \neq j$, and $\delta_{ii} = 1$. This is the *identity* $n \times n$ matrix. Its determinant is $\det(I) = 1$.

4. If two rows of M are equal, then $\det(M) = 0$
 5. Define the *transpose* of M thus: Let $M = (a_{ij})$, define $b_{ij} = a_{ji}$; the transpose $M^t = (b_{ij})$. Then $\det(M^t) = \det(M)$.
 6. If P and Q differ only in their i th rows, and if M is the matrix obtained from P by replacing the i th row by the sum of those in P and in Q , then $\det(M) = \det(P) + \det(Q)$.
 7. If P is equal to Q except that the i th row of P is a constant multiple (by $\lambda \in A$) of the i th row of Q , then $\det(P) = \lambda \cdot \det(Q)$.
 8. The *adjoint* of P is the matrix P^{adj} described thus: Strike out the i th row and j th column of P , obtaining the $(n-1) \times (n-1)$ matrix P_{ij} . Define $q_{ij} = (-1)^{i+j} \det(P_{ij})$. Then P^{adj} is the matrix (q_{ij}) . What then happens is that $P^{\text{adj}} \cdot P = D$ where $D = (d_{ij})$, with $d_{ij} = 0$ for $i \neq j$, and $d_{ii} = \det(P)$. In other words, $P^{\text{adj}} \cdot P$ is a diagonal matrix, each diagonal term being $\det(P)$.
 9. If $Q \cdot P = I$, then $\det(P)$ is a unit in the comring A . Conversely, if $\det(P) = u$ is a unit, then such an inverse Q can be found by multiplying all entries in P^{adj} by u^{-1} .
 10. If $Q \cdot P = I$, then $P \cdot Q = I$.
 11. If $\varphi: A_1 \rightarrow A_2$ is a comring-map, and if P is an $n \times n$ matrix over A_1 , then $\varphi(P)$ denotes the $n \times n$ matrix over A_2 obtained by applying φ to each entry of A_1 . Then $\det(\varphi(P)) = \varphi(\det(P))$.
- Etc., etc.

Here are a couple of other results that will be useful. First, let us define a *monic* polynomial in $A[x]$. This means a polynomial whose leading coefficient is 1.

8.1. Theorem. *Suppose A is a comring, I the identity $n \times n$ matrix, and K an A -ideal. Suppose M is an $n \times n$ matrix all of whose entries are in K . Then $\det(I - M) = 1 + \mu$, where $\mu \in K$.*

Proof: Let $B = A/K$ and $\varphi: A \rightarrow B$ the natural comring-map. Then $\varphi(I - M) = I$ the identity matrix over B . Using property (11) above,

$$\varphi(1) = 1 = \det(\varphi(I - M)) = \varphi(\det(I - M)).$$

Thus, $\det(I - M) - 1 \in K$. ■

8.2. Theorem. *If M is any $n \times n$ matrix over the comring A , then we can consider determinants of matrices over $A[x]$. The conclusion is that the characteristic polynomial of M , that is, $\det(xI - M)$, is a monic polynomial of degree n . ■*

* EXERCISE 8.1: Prove it.

9. The Krull intersection theorem

• **9.1. Product of ideals.** Suppose that I and J are ideals in the comring A . Then, by $I \cdot J$ we mean the ideal generated by all products xy , for $x \in I$ and $y \in J$. For a positive integer n , we define $I^1 = I$ and $I^{n+1} = I \cdot I^n$. And we define $I^\infty = \bigcap_1^\infty I^n$.

9.2. Krull Intersection Theorem. *If A is a Noetherian ring, and I and J are ideals in A such that $J \subset I^\infty$, then $I \cdot J = J$.*

Proof: The problem is to show that every element α of J belongs to the product $I \cdot J$.

We note that I is generated as an A -ideal, by a finite set $\{b_1, \dots, b_n\}$. We first consider

$$A[x] = A + Ax + Ax^2 + \dots + Ax^n + \dots.$$

This contains the subring

$$B = A + Ix + I^2x^2 + \dots + I^n x^n + \dots$$

and the set (an ideal both in $A[x]$ and in B):

$$\tilde{J} = J + Jx + Jx^2 + \dots + Jx^n + \dots.$$

Note that $\tilde{J} \subset B$ since $J \subset I^\infty$. There is a ring-map $A[y_1, \dots, y_n] \rightarrow B$ taking y_i to $b_i x$. The nature of B shows that this map is surjective; by the Hilbert basis theorem 4.1, $A[y_1, \dots, y_n]$ is Noetherian, and hence B (being a quotient ring of $A[y_1, \dots, y_n]$) is Noetherian.

Now, \tilde{J} is an ideal in B , and hence is finitely generated as a B -ideal, by polynomials $\{f_1, \dots, f_k\}$ which belong to \tilde{J} ; these generators are all polynomials in x of degree bounded by some integer m . Consider the element αx^{m+1} in \tilde{J} , where $\alpha \in J$. This element must be a sum of terms of the form $g(x)f_i(x)$, where $g \in B$, and where $f_i \in \tilde{J}$ has degree less than $m + 1$. The coefficient of x^{m+1} in such a term is a sum of products of the coefficient of some *positive* power of x in g (all these coefficients are in I) with a coefficient of a power of x in f_i ; it thus belongs to $I \cdot J$. Hence $\alpha \in I \cdot J$. ■

Here is a corollary of some interest:

9.3. Corollary. *If A is a Noetherian ring, then the intersection of all powers of all maximal ideals of A is $\{0\}$.*

Proof: Given $a \in A$ with $a \neq 0$, define (the “annihilator” of a):

$$I_a = \{x \in A \mid xa = 0\}.$$

Then I_a is an ideal in A , and $1 \notin I_a$. There is a maximal ideal M containing I_a . The claim is that there is some $n > 0$ such that $a \notin M^n$. Otherwise $\langle a \rangle \subset M^\infty$, and by the Krull Intersection Theorem, $M \cdot \langle a \rangle = \langle a \rangle$. Thus there exists $m \in M$ such that $ma = a$, or $(1 - m)a = 0$. That is, $1 - m \in I_a \subset M$. Hence we would have $1 \in M$. ■

Here is an interesting fact, peripherally related to this topic:

9.4. Theorem. *Suppose J and K are ideals in the comring A , and J is finitely generated, and $K \cdot J = J$. Then there is an element $\kappa \in K$, such that for all $x \in J$, $\kappa x = x$.*

Proof: Let $\{b_1, \dots, b_n\}$ be a finite set of generators of J as A -ideal. Since $K \cdot J = J$, we can write

$$b_i = \sum_j c_{ij} b_j, \text{ where } c_{ij} \in K.$$

Let $C = (c_{ij})$ be this $n \times n$ matrix, and let I be the $n \times n$ identity matrix. We regard (b_i) as an $n \times 1$ (column) matrix. Then we have

$$(*) \quad (I - C)(b_i) = 0.$$

Multiply this on the left by the adjoint of $I - C$, and we see that, for all i , where $\delta = \det(I - C)$,

$$\delta b_i = 0,$$

and hence $\delta x = 0$ for all $x \in J$. Now, by properties of determinants noted previously (8.1), since all the entries of C belong to K , we can define $\kappa = 1 - \delta$, and we then know that $\kappa \in K$, and that $\kappa x = x$ for all $x \in J$. ■

The “ $A+V$ ” construction yields this corollary of the Krull intersection theorem:

9.5. Krull Intersection Theorem for Modules. *If A is a Noetherian ring, V a finitely generated module over A , I an ideal of A , and if S is a submodule of V such that $S \subset I^\infty \cdot V$, then $S = I \cdot S$. ■*

Furthermore, by the theorem just proved by means of determinants, there is $\kappa \in I$ such that for all $x \in S$, it is true that $\kappa x = x$.

10. A Field, FG as a Ring, is Finite; short proof

In 1940, in *Matem. Sbornik* **8(50)**, pages 406–421, A. Malcev had a paper called (my translation) “On Faithful Representations of Infinite Groups by Matrices.” Among other things, Malcev, in this paper, proved that a finitely generated group of matrices, of some finite size over a commutative field, is residually finite. In the proof of this, he essentially shows that a field which is finitely generated as a ring is, in fact, finite. This is a special case of a theorem of Goldman and Krull, generalizing Hilbert’s “Nullstellensatz.”

This section will be devoted to a discussion of this theorem that a finitely generated comring which is a field is finite. I shall say a little about Malcev’s idea (which I haven’t figured out much yet); and then I shall quote a letter from Peter Shalen, in which he gives a simple, one-page proof. If you add to this the Krull Intersection Theorem (which is an easy consequence of the Hilbert Basis Theorem), then it is not hard to prove that every finitely generated comring is residually finite; from which it follows easily that finitely generated groups of matrices over any comring (not necessarily a field) are residually finite.

Later, in the rest of this chapter of the notes, I present a proof of the theorem of Goldman and Krull in greater generality.

• **10.1. Malcev’s Lemma 3.** First, a word about Malcev’s proof itself. A crucial point is Lemma 3 in his paper, which, together with the remarks preceding it (from a brief glance at the paper, it seems that no proof is given), goes as follows (my translation again):

“An application of the notion of group-limit [M. thinks of residually finite groups as limits of finite groups] to the study of representations is based on an observation from number theory, which for clarity we formulate as a separate lemma.

Lemma 3. *Suppose there is a system consisting of a finite number of equations of the form $f_\alpha(x_1, \dots, x_k) = 0$ and a finite number of inequalities $f_\alpha(x_1, \dots, x_k) \neq 0$, where the left sides are integral polynomials in the variables x_1, \dots, x_k . If this system is solvable in some field of characteristic zero, then it is solvable in an infinite number of prime fields of finite characteristic.”*

In other words, the equations define an ideal I in $\mathbf{Z}[x_1, \dots, x_k]$, and the quotient of this ideal we can suppose (for the purposes of this Lemma) to be an integral domain of characteristic zero. Then Malcev says that the ideal I is the intersection of ideals J where the quotient by J is a field of the form \mathbf{Z}_p for p a prime. This is

more than enough to imply that if M is any maximal ideal in $\mathbf{Z}[x_1, \dots, x_k]$, then the factor field has finite characteristic.

[I have heard that this Lemma 3 of Malcev was a well-known “folk theorem.”]

On the next page, in the proof of Theorem VIII, Malcev considers the case of finite characteristic and shows, essentially, that such a field which is a factor ring of $\mathbf{Z}[x_1, \dots, x_k]$ is in fact finite. His proof is based on a little bit of the theory of algebraic varieties over an algebraically closed field of finite characteristic.

• **10.2. Shalen’s proof.** Peter Shalen has given me a proof (in a letter dated Feb. 15, 1990) that every (commutative) field which is finitely generated as a ring is a finite field. Here (slightly edited) it is:

10.3. Lemma. *If E is a finite extension of the field F , and if E is finitely generated as a ring, then so is F .*

Proof: We can use induction on the degree of the extension to reduce to the case where E is a simple extension $F[x]$. To prove it in this case we look at the minimal polynomial $f(X) = X^d + a_{d-1}X^{d-1} + \dots + a_0$ of x , and we let y_1, \dots, y_m be a finite generating set for E as a ring. Then each y_i can be written in the form $\sum_{j=0}^{d-1} b_{ij}x^j$ for some $b_{ij} \in F$. Let R be the sub-ring of F generated by all the a_j and b_{ij} . Given any element $z \in F$, we can write z as an integer polynomial in the y_i , and hence as a polynomial in x with coefficients in R . Since the a_i are in R , we can then rewrite z as a polynomial of degree $< d$ in x with coefficients in R , say $z = \sum_{j=0}^{d-1} r_j x^j$. But since z and the r_j belong to F , and $1, x, \dots, x^{d-1}$ form a basis of E as an F -vector space, we must have $r_0 = z$ (and $r_j = 0$ for $j > 0$). So $z \in R$. This shows that $R = F$, so that F is finitely generated. ■

10.4. Theorem. *If E is a field which is finitely generated as a ring, then E is finite.*

Proof: E is a finitely generated extension of its prime field P . So E has a finite transcendence basis x_1, \dots, x_n , and E is a finite extension of $F = P(x_1, \dots, x_n)$. By the Lemma, F is finitely generated as a ring.

Suppose first that $n = 0$. Then $F = P$. If P has prime characteristic then $F = P$ is finite, and hence so is E . If P has characteristic zero then F is isomorphic to \mathbf{Q} ; but \mathbf{Q} is not finitely generated as a ring because there are infinitely many primes in \mathbf{Z} .

Now suppose that $n > 0$. Then E is isomorphic to $K(X)$, where K is a rational function field in $n - 1$ indeterminates over P . But again, if K is any field, the rational function field $K(X)$ cannot be finitely generated as a ring: This follows from the fact that $K[X]$ has infinitely many (monic) irreducible polynomials, which is proved in the same way as the corresponding fact for \mathbf{Z} . ■

We now derive some consequences of this fact.

10.5. Theorem. *Suppose that D is an integral domain which is finitely generated as a ring. Suppose that $a \in D$ and $a \neq 0$. Then there is a maximal ideal M of D which does not contain a . The field D/M is finite. [In other words, finitely generated integral domains are “residually finite fields.”]*

Proof: Let K be the field of fractions of D . Let $D[x]$ be the polynomial ring and let $\varphi: D[x] \rightarrow K$ be given by $d \mapsto d$ and $x \mapsto 1/a$. Let $H = \varphi(D[x])$. In other words, H consists of all fractions whose denominators are powers of a . We have $D \subset H$, and D is generated by D plus the additional element $1/a$. Thus, H is finitely generated as a ring. There exists a maximal ideal M_1 of H . Since a is a unit in H , we have $a \notin M_1$. By 10.4, H/M_1 is a finite field. Now, define $M = D \cap M_1$. This is clearly an ideal in D , and we have $D/M = D/(D \cap M_1) \subset H/M_1$. Thus, D/M is a finite integral domain and so, by Exercise 2.3, is a finite field. And so M is a maximal ideal in D . ■

In fact, we can extend this result to comrings rather than domains, using the Krull Intersection Theorem, in the following way:

By a *finitely generated comring*, one means a ring A isomorphic to $\mathbf{Z}[x_1, \dots, x_n]$ modulo some ideal. The Hilbert Basis Theorem 4.1 implies that $\mathbf{Z}[x_1, \dots, x_n]$ is Noetherian, and so its quotient A is also Noetherian.

10.6. Theorem (Finitely generated comrings are residually finite). *If A is any finitely generated comring, and if $a \in A \setminus \{0\}$, then there is an ideal J of A such that $a \in A \setminus J$ and A/J is a finite ring.*

Proof: By Corollary 9.3 of the Krull Intersection Theorem, there is a maximal ideal M of A and an integer $n \geq 1$, such that $a \in A \setminus M^n$. The claim is that A/M^n is finite. For $n = 1$, this is implied by Theorem B, since the inverse image of M in $\mathbf{Z}[x_1, \dots, x_k]$ is of “finite index.” Now, M^n is an ideal in the Noetherian ring A , hence generated by finitely many elements; the quotient M^n/M^{n+1} has the

structure of an (A/M) -module; it is thus a finite-dimensional vector space over the finite field A/M , hence a finite set. The number of elements $|X|$ of X satisfies

$$\left| \frac{A}{M^{n+1}} \right| = \left| \frac{A}{M^n} \right| \cdot \left| \frac{M^n}{M^{n+1}} \right|. \blacksquare$$

• **10.7. Residually finite group.** We call a group G “residually finite” when, for each $g \in G \setminus \{1\}$, there is a normal subgroup N of G such that $g \notin N$ and G/N is finite. Clearly, every subgroup of a residually finite group is residually finite.

Given some positive integer n , we define $GL_n(A)$ to be the group of invertible $n \times n$ matrices over the comring A .

10.8. Theorem. (a) *If A is a finitely generated comring, then $GL_n(A)$ is residually finite.*

(b) *If G is a finitely generated subgroup of $GL_n(B)$, where B is a comring, then G is residually finite.*

Proof: (a) Suppose $\alpha \in GL_n(A)$ and $\alpha \neq I$; then the matrix α differs from the identity matrix I at some entry. By **10.6**, A has an ideal J , such that A/J is finite and such that, modulo J , it is still true that $\alpha \neq I$. We can use the quotient homomorphism $\varphi: A \rightarrow A/J$ to define a homomorphism $\tilde{\varphi}: GL_n(A) \rightarrow GL_n(A/J)$ in which $\tilde{\varphi}(\alpha) \neq I$. Now, $GL_n(A/J)$ is a finite group; hence $N = \ker \tilde{\varphi}$ is a normal subgroup of finite index in $GL_n(A)$ with $\alpha \notin N$.

(b) Suppose G is generated by the matrices $\{g_1, \dots, g_k\}$. Let A be the subring of B generated by all the entries in the matrices

$$\{g_1, \dots, g_k, g_1^{-1}, \dots, g_k^{-1}\}.$$

Then A is a finitely generated comring. The homomorphism

$$GL_n(A) \rightarrow GL_n(B)$$

is injective; its image contains G . Hence G is isomorphic to a subgroup of $GL_n(A)$, which is residually finite by part (a). \blacksquare

The above theorem, for matrices over fields rather than comrings, is one of the main results of the paper by Malcev referred to earlier.

▷ **EXERCISE 10.1:** Free groups are residually finite. Free monoids are residually finite (define “residually finite” for monoids).

* EXERCISE 10.2: Is it possible to have an integer $n \geq 2$, such that $2^n \equiv 1 \pmod{n}$? Show that n has to be odd; that n cannot be divisible by 3.

In general, if p is the smallest prime divisor of $n = pq$, then, since $2^p \equiv 2 \pmod{p}$, we would have $2^q \equiv 2^{pq} \equiv 1$. Hence, q is divisible by the order of the element 2 in the multiplicative group of units in \mathbf{Z}_p ; this order divides $p - 1$.

* EXERCISE 10.3: Consider the group G with presentation

$$\langle a, x \mid axa^{-1}xax^{-1}a^{-1} = x^2 \rangle.$$

[This is an example due to Gilbert Baumslag.] If we map this group into a finite group, and the order of the image y of x is n , then, since

$$(axa^{-1})^k x (axa^{-1})^{-k} = x^{2^k},$$

the image y will have the property that $y^{2^n} = y$. Apply the preceding exercise to show that y is the identity element.

Conclude that if X is an invertible matrix of some finite size over some commutative ring, and if we assume that X is conjugate to X^2 and that the conjugating matrix is itself conjugate to X , then X is the identity matrix.

* EXERCISE 10.4: Similar to the above exercise, investigate the group defined by G. Higman

$$\langle a, b, c, d \mid aba^{-1} = b^2, bcb^{-1} = c^2, cdc^{-1} = d^2, dad^{-1} = a^2 \rangle.$$

Every finite quotient group of this group is trivial, and hence every map of this group into a group of matrices over a comring is degenerate.

** EXERCISE 10.5: Prove, by one means or other, that Higman's group is non-trivial; and that the element x in Baumslag's group is not the identity element. Both of these groups are related to the group $H =$

$$\langle a, b \mid aba^{-1} = b^2 \rangle,$$

which can be embedded in a matrix group:

$$a \mapsto \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad b \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

To say something about Baumslag's group, we would like to embed H in a bigger group in which a is conjugate to b . Since both a and b have infinite order, this can be done by the HNN-construction. Similarly, Higman's group can be obtained from H by doing a few amalgamated free products. It is possible to do all this kind of thing in a logically more primitive way, representing the groups into permutation groups (of infinite sets) in clever ways.

11. Factorial domains

• **11.1. Irreducible elements and units.** Let D be an integral domain. There are three kinds of elements: The element 0; the units, those $x \in D$ for which there exists $y \in D$ such that $xy = 1$; and "other."

Let us consider the "other" type. An element p of this sort is called *irreducible* when: $\forall b, c \in A$, if $p = bc$, then either b or c is a unit.

Two non-zero elements a and b are said to be *associate* when there is a unit u of A such that $a = bu$. In other words, the group of units acts on the non-zero elements, by multiplication; the set of units forms one orbit; the orbits are the "associateness classes."

* EXERCISE 11.1: When $a = bc$ and b and c are both "other," we have, on the principal ideals, $\langle a \rangle \subset \langle b \rangle$ and $\langle a \rangle \neq \langle b \rangle$. If $\langle a \rangle = \langle b \rangle$, then a and b are associate, and if, additionally, $a = bc$, then c is a unit. If a is "other" and a cannot be written as a product of irreducible elements, then there is an infinite, properly ascending chain of principal ideals, starting with $\langle a \rangle$.

When $a = bc$, we say " b divides a ," " b is a divisor of a ," " a is a multiple of b ," " $b \mid a$," or " $\langle a \rangle \subset \langle b \rangle$."

• **11.2. Factorial domain.** Now here is the definition: A *factorial domain* is an integral domain D such that:

1. There is no infinite, properly ascending chain of principal ideals:

$$\langle a_1 \rangle \subset \langle a_2 \rangle \subset \cdots \subset \langle a_n \rangle \subset \cdots$$

with each $\langle a_i \rangle \neq \langle a_{i+1} \rangle$.

2. For every irreducible p in D , the ideal $\langle p \rangle$ is a prime ideal.

Equivalently:

1'. Every "other" element of D can be written as a product of some finite number of irreducible elements.

2'. Two such expressions

$$a = p_1 p_2 \cdots p_n = q_1 \cdots q_m$$

for p_i, q_i irreducible, have $m = n$, and, up to a reordering of the factors, each p_i is associate to q_i .

These are related by implications $(1) \implies (1')$, $(2) \implies (2')$, plus the fact that $(1')$ and $(2') \implies (1)$ and (2) . As pointed out earlier, the first version of (1) implies the second. The first version of (2) says that if p is irreducible and $p \mid ab$, then either a or b is a multiple of p . Thus, in trying to deduce the second version of (2) from the first version, we have $p_1 \mid q_1 \cdots q_m$ and hence p_1 divides some q_i ; since q_i is irreducible, then p_1 and q_i are associate; they can be cancelled from the equation since D is a domain; the second version of (2) follows by induction.

* EXERCISE 11.2: To show that the conjunction of the second versions imply the first is left to the reader.

Because of this, a factorial domain is also called a "unique factorization domain" or UFD.

11.3. Theorem. *If D is a principal ideal domain, then D is factorial.*

Proof: The fact is that every Noetherian domain obviously satisfies (1). The problem is (2). Let p be an irreducible element of the p.i.d. D ; the claim is that $\langle p \rangle$ is a maximal ideal, and so, *a fortiori*, prime: If $\langle p \rangle$ is contained in an ideal $I \neq A$, then I is principal, $I = \langle a \rangle$; and $p \in \langle a \rangle$ says $a \mid p$; and thus, since p is irreducible, either a is a unit or associate to p ; but a cannot be a unit, since $I = \langle a \rangle \neq A$; therefore, $\langle a \rangle = \langle p \rangle$. ■

In a factorial domain, if we consider two elements a and b , these two elements have a *greatest common divisor* $c = \gcd(a, b)$, which is unique up to associateness. This element c has the property that $c \mid a$ and $c \mid b$, and the property that if x is any element such that $x \mid a$ and $x \mid b$, then $x \mid c$. One finds c by writing a and b as products of irreducible factors, selecting all the common factors (up to associateness), and taking their product. A unit such as 1 divides everything, and everything divides 0.

In a principal ideal domain, $\gcd(a, b)$ can be found thus: $\langle a, b \rangle$, the ideal generated by a and b , is principal: $\langle a, b \rangle = \langle c \rangle$, and $c = \gcd(a, b)$. In a Euclidean domain,

the gcd of a and b can be found by a little computer program (the “Euclidean algorithm”):

{ If $a = 0$, then $\gcd(a, b) = b$. [Stop]
 If $a \neq 0$, divide b by a :
 $b = aq + r$ [with $e(r) < e(a)$]
[Note: $\langle a, b \rangle = \langle r, a \rangle$.]
 Set $b = a$.
 Set $a = r$.
 } repeat.

12. Modules over a principal ideal domain

This is a subject which I am only going to sketch. A finitely generated module over a ring A is A -module-equivalent to A^n/S , where S is some submodule of A^n and n is some non-negative integer. If A is a Noetherian ring, we can find a finite set of generators for S , and thus have all the information about A^n/S in the form of a $k \times n$ matrix $M = (a_{ij})$ over A ; we can think of S as the “row space” of this matrix. We can get other generators for S by multiplying M on the left by an invertible $k \times k$ matrix P , getting PM . We can change the way we look at A^n , change its basis, by multiplying on the right by an invertible $n \times n$ A -matrix Q . Thus, to find out something about A^n/S , we can manipulate M to PMQ and see what happens.

In case D is a principal ideal domain, a submodule of D^n can be generated by n or fewer elements. The matrix that results can be padded with 0-rows to make it square. This $n \times n$ matrix M can then be changed to PMQ , where P and Q are invertible $n \times n$ D -matrices, so that the resulting matrix $M' = PMQ$ is diagonal; the non-zero terms are only on the diagonal: d_1, d_2, \dots, d_n . They are arranged so that $d_i \mid d_{i+1}$. The rows with $d_i = 1$ and their corresponding columns can be removed without changing A^n/S . The result we get is this:

12.1. Theorem. *Let V be a finitely generated module over a principal ideal domain D . Then there are ideals $\langle d_1 \rangle, \dots, \langle d_n \rangle$ in D , such that*

$$D \neq \langle d_1 \rangle \supset \langle d_2 \rangle \supset \cdots \supset \langle d_n \rangle,$$

and such that V is D -module equivalent to

$$(D/\langle d_1 \rangle) \times (D/\langle d_2 \rangle) \times \cdots \times (D/\langle d_n \rangle).$$

This sequence of ideals is uniquely determined by V .

Sketch of proof: We manipulate the matrix M by doing row and column operations. If D is a Euclidean domain, this suffices. Otherwise, we can obtain greatest common divisors of terms of M by using 2×2 invertible matrices. Thus we read off the structure of $X \approx D^n/S$ from the resulting diagonal matrix.

The uniqueness of the structure requires some discussion along the following lines:

We say that V is *torsion-free* when, for all $a \in D$ and all $x \in V$, if $ax = 0$, then $a = 0$ or $x = 0$. It follows that every finitely generated torsion-free module over a principal ideal D is of the form D^k . (The example of the abelian group, the additive group of the rational numbers, shows that this is not true without the “finitely generated” assumption.)

Given any D -module V , we define the *torsion submodule* $T(V)$ to be the set

$$\{x \in V \mid \exists a \in A \ni a \neq 0 \text{ and } ax = 0\}.$$

It further follows from the matrix argument that a finitely generated module over the principal ideal domain D is of the form $T(V) \times D^k$, where this D^k can be identified with the torsion-free quotient module $V/T(V)$. It is not generally true, without the finitely generated assumption, that $T(V)$ is a direct factor of V .

Now, if $V = T(V)$, that is, if V is a torsion module, we can look at the annihilators of the elements of V , and from our matrix discussion it follows that d_1 generates the largest annihilator of a non-zero element x ; this shows that d_1 is uniquely determined by V itself, up to associateness. It can be shown, with more argument (omitted), that the other d_i are also determined by V . ■

• **12.2. Abelian groups.** For an abelian group G , as module over \mathbf{Z} , if $G = T(G)$, we call G a torsion abelian group. In particular, if G is a finitely generated, torsion abelian group, then the number of elements of G , which we call $|G|$, is the product $d_1 \cdot d_2 \cdots d_n$.

Concerning finitely generated abelian groups (as \mathbf{Z} -modules), we use the shorthand \mathbf{Z}_n for $\mathbf{Z}/\langle n \rangle$, and we can, for instance, describe all abelian groups of order 32 as:

$$\begin{array}{ll} \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_2 & \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_4 \\ \mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_8 & \mathbf{Z}_2 \times \mathbf{Z}_{16} \end{array}$$

$$\mathbf{Z}_{32} \quad \mathbf{Z}_2 \times \mathbf{Z}_4 \times \mathbf{Z}_4 \quad \mathbf{Z}_4 \times \mathbf{Z}_8$$

* EXERCISE 12.1: What are all the abelian groups of orders 1980, 1984, 1986, 1987, 1988, 1989, 1990, 2000? [There are 15 such groups of order 2000, but only one of order 1990.]

One corollary of this is:

12.3. Theorem. *Every finite subgroup of the multiplicative group of a field is cyclic.*

Proof: The reason is that, in a field F , the equation $x^n = 1$ can have at most n solutions; but if, in the description of this group as a direct product of cyclic groups, each of whose order divides the next, there is more than one factor, then there are too many solutions to such an equation. ■

* EXERCISE 12.2: This is particularly interesting for finite fields. For instance, the multiplicative group of \mathbf{Z}_{37} is cyclic of order 36, instead of being one of the other three possible abelian groups of this order; a generator is easily found, such as 2. The multiplicative group of \mathbf{Z}_{71} is cyclic of order 70; all abelian groups of order 70 are cyclic, but it is harder to find a generator of this group; the smallest one is 7. There is a field of order 16, which can be described as

$$\mathbf{Z}_2[x]/\langle x^4 + x^3 + x^2 + x + 1 \rangle.$$

Its multiplicative group is cyclic of order 15; the coset containing x has multiplicative order 5.

13. D factorial $\implies D[x]$ factorial

In the early part of his book *Disquisitiones Arithmeticae* (1801), Gauss proved what amounted to the following theorem:

13.1. Theorem. $\mathbf{Z}[x]$ is a factorial domain.

Proof: (1) Since \mathbf{Z} is a domain, $\mathbf{Z}[x]$ is a domain. The degree of $f(x) \cdot g(x)$ is the sum of the degrees of f and g . A unit in $\mathbf{Z}[x]$ is the same as a unit in \mathbf{Z} .

(2) Every irreducible element (= prime number) in \mathbf{Z} is irreducible in $\mathbf{Z}[x]$. From these facts, it follows that:

(3) Every non-zero, non-unit in $\mathbf{Z}[x]$ can be expressed in some way as a product of irreducible elements of $\mathbf{Z}[x]$.

(4) A polynomial $f \in \mathbf{Z}[x]$ of degree ≥ 1 is irreducible in $\mathbf{Z}[x]$ if and only if (a) the greatest common divisor of its coefficients is 1, and (b) as a polynomial in $\mathbf{Q}[x]$, it is irreducible.

—The crucial point in proving (4) is this:

(5) If $f \in \mathbf{Z}[x]$, $g, h \in \mathbf{Q}[x]$, $f = gh$, where g and h are polynomials with rational coefficients, of degrees k and l respectively, with both k and l greater than 0, then f can be written as a product of polynomials with integer coefficients of the same degrees.

To see this, we can put all the coefficients of g over a common denominator, and similarly for h :

$$g(x) = \frac{a}{b}\tilde{g}(x) \quad h(x) = \frac{c}{d}\tilde{h}(x)$$

where a, b, c, d are non-zero integers, $\tilde{g}(x), \tilde{h}(x) \in \mathbf{Z}[x]$; and where the gcd of the coefficients of $\tilde{g}(x)$ is 1, and the gcd of the coefficients of $\tilde{h}(x)$ is 1. Thus:

$$f(x) = \frac{ac}{bd}\tilde{g}(x)\tilde{h}(x) = \frac{\alpha}{\beta}\tilde{g}(x)\tilde{h}(x),$$

where α and β are integers without any common factor.

Now suppose we could prove, “If $\tilde{g}(x), \tilde{h}(x) \in \mathbf{Z}[x]$ are polynomials, for each of which the gcd of the coefficients is 1, then the gcd of the coefficients of $\tilde{g}(x)\tilde{h}(x)$ is 1.” Then, by examining $\beta f(x) = \alpha\tilde{g}(x)\tilde{h}(x)$, factoring out a common integer factor from $f(x)$ if possible, we would see that α is a multiple of β . And hence,

$$f(x) = \left(\frac{\alpha}{\beta}\tilde{g}(x)\right)\tilde{h}(x)$$

is the required $\mathbf{Z}[x]$ -factorization.

(5') A comment: If $f(x) = g(x)\cdot h(x)$, where $f, h \in \mathbf{Z}[x]$, $g \in \mathbf{Q}[x]$, and if the greatest common divisor of the coefficients of h is 1, then it also follows from the above argument that g has integral coefficients.

Thus, the basic problem in proving (5) is the following:

(6) [“Gauss’s Lemma,” §42 in *Disq. Arith.*] If $g(x), h(x) \in \mathbf{Z}[x]$ are such that the greatest common divisor of the coefficients of $g(x)$ and the gcd of the coefficients of $h(x)$ are both equal to 1, then the polynomial $f(x) = g(x)\cdot h(x)$ has the same property, the gcd of the coefficients of f is 1.

Proof: If not, there is some prime number p such that all coefficients of $f(x)$ are divisible by p . We then look at what happens mod p . That is, we examine the

homomorphism $\varphi: \mathbf{Z}[x] \rightarrow \mathbf{Z}_p[x]$. We note that \mathbf{Z}_p is a domain, and hence $\mathbf{Z}_p[x]$ is a domain. Let $F = \varphi(f)$, $G = \varphi(g)$, $H = \varphi(h)$. Then $F = G \cdot H$ and, furthermore, $F = 0$. Therefore, either $G = 0$ or $H = 0$. If $G = 0$, this is tantamount to saying that p divides all the coefficients of $g(x)$, contrary to hypothesis.

(7) It remains to say why each irreducible element p of $\mathbf{Z}[x]$ generates a prime ideal. If p is of degree 0, then p is simply a prime number, and the quotient $\mathbf{Z}[x]/(\mathbf{Z}[x]\langle p \rangle) = \mathbf{Z}_p[x]$ is a domain.

If p has degree > 0 , then $p(x)$, by (6), is irreducible in $\mathbf{Q}[x]$. Since \mathbf{Q} is a field, $\mathbf{Q}[x]$ is a Euclidean domain, and so $\mathbf{Q}[x]/(\mathbf{Q}[x]\langle p \rangle)$ is a field. We examine the inclusion $\mathbf{Z}[x] \rightarrow \mathbf{Q}[x]$ and claim

$$(\mathbf{Q}[x]\langle p \rangle) \cap (\mathbf{Z}[x]) = \mathbf{Z}[x]\langle p \rangle.$$

This means: If there is a rational polynomial $g(x)$ with the property that the product $f(x) = g(x) \cdot p(x)$ belongs to $\mathbf{Z}[x]$, then $g(x)$ also belongs to $\mathbf{Z}[x]$. This has been remarked upon in (5'); this uses the fact that, since $p(x)$ is irreducible in $\mathbf{Z}[x]$ and of degree > 0 , the greatest common divisor of the coefficients of $p(x)$ is 1.

Thus, the inclusion induces an injective map of comrings

$$\mathbf{Z}[x]/(\mathbf{Z}[x]\langle p \rangle) \rightarrow \mathbf{Q}[x]/(\mathbf{Q}[x]\langle p \rangle).$$

Thus $\mathbf{Z}[x]/\langle p \rangle$, a subring of a field, is an integral domain; hence $\langle p \rangle$ is a prime ideal in $\mathbf{Z}[x]$. ■

The above theorem immediately generalizes:

13.2. Theorem. *If D is a factorial domain, then $D[x]$ is a factorial domain.*

Proof: The properties of \mathbf{Z} which are used in the proof are exactly the properties used in the definition of “factorial domain.” Note in particular that in proving (3), we did not use the fact (which *could* have been used to prove it) that \mathbf{Z} is Noetherian. In (4) and (5), we used the rational numbers \mathbf{Q} ; this can be replaced, without changing anything essential, by the notion of the *field F of fractions of the domain D* . A “prime number” p becomes, in this context, an “irreducible element” of D ; its basic property, used in (6) and (7), is that $\langle p \rangle_D$ is a prime ideal in D ; this is a fundamental fact about factorial domains. ■

14. The Complex Numbers

[This historical material is derived from the books *Episodes from the Early History of Mathematics* by A. Aaboe, and *A Source Book in Mathematics 1200–1800* by D. J. Struik.]

The Babylonians developed quite a lot of arithmetic; most of what we know about this comes from miscellaneous clay tablets dating from around 1700 b.c.; I don't believe that any textbook has survived, only things in the nature of school exercises and a few tables. Their numerical notation (positional, base 60) is so obvious that even modern Babylonian scholars could decipher it. (The Mayans had an even better notation, base 20, which included zero, unknown in Babylonia in 1700 b.c.; but almost nothing is known of how well they developed arithmetic.) We know that they had a very accurate approximation of $\sqrt{2}$, namely

$$1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3},$$

and that they had a way to find triples of integers (x, y, z) such that $x^2 + y^2 = z^2$, one of which was $(12709, 13500, 18541)$. They had a general method for solving quadratic equations posed in the form $x + y = a$, $xy = b$; and some of their homework exercises of this sort do not have real solutions. Hence, one can make the purely imaginary conjecture that they knew something about the complex numbers.

• **14.1. Cardan.** The first written record of square roots of negative numbers occurs in the book *Ars Magna* by Cardano (1545). Cardano gives in this book a general method for solving cubic and quartic equations (due to his contemporaries del Ferro, Tartaglia, and Ferrari); but he goes to great trouble only to do examples involving positive numbers and roots of positive numbers. The general method described by Cardano, applied to $x^3 - 3x + 1 = 0$, would solve it by substituting $x = y + y^{-1}$, obtaining

$$y = \sqrt[3]{\frac{-1 + \sqrt{-3}}{2}}.$$

Perhaps a more peculiar example is the equation

$$x^3 = 15x + 4,$$

having a nice positive root $x = 4$, but for which Cardan's formula gives the answer

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Cardano does not write down *these* answers. But, at one point Cardano considers explicitly the pair of equations, $x + y = 10$, $xy = 40$, and writes down the answer as “5p: Rx m: 15” and “5m: Rx m: 15,” or, in modern notation, “ $5 \pm \sqrt{-15}$.” He makes comments about this such as, “tortures notwithstanding,” and “this is as subtle as it is useless.”

• **14.2. Conjecture on existence of roots of polynomials.** Later in the 1500s and early 1600s, mathematicians got used to the idea of square roots of negative numbers; they also developed modern notation and started using letters such as “A, B,” for arbitrary constant quantities, instead of simply giving example after example. By 1600, the conjecture was around that every real polynomial has a complex root; this was explicitly formulated in print by Girard (1629). After this, many well-known mathematicians attempted to prove this, including D’Alembert (1748), Euler (1749), De Foncenet (1759), Lagrange (1772).

A real understanding of what was going on was first achieved by Gauss in his doctoral dissertation in 1799. Gauss’s idea (the one in his dissertation; later he developed several other proofs of the “fundamental theorem of algebra”) went like this: Suppose that $f(x)$ is a monic, real polynomial of degree n . Let A be the set of complex numbers z where the real part of $f(z)$ is zero; let B be the set where the imaginary part is zero. On a large circle C in the complex plane, f behaves like z^n , and so A will intersect this circle in $2n$ points, and B will intersect it in $2n$ points which alternate with the intersections of A . Gauss then claims that the curves A and B must intersect somewhere within the disk bounded by C ; the zeros of f consist of the set $A \cap B$. — The modern criticism of this might go: The fact that $A \cap C$ consists of $2n$ points is proved by an argument which uses the intermediate-value property of continuous real functions, a theorem generally credited to Bolzano; a reasonably correct proof of this had to wait until the nature of the real numbers was understood (by Weierstrass and others). More seriously, the fact $A \cap B \neq \emptyset$ was based on an argument using something like the Jordan Curve Theorem (first proved rigorously by O. Veblen around 1905) and assumptions about the nature of A and B which follow from differential geometry as developed later by Gauss himself. In a way, what Gauss was doing was to consider two 2-spheres E_1 and E_2 in $S^2 \times S^2$, where E_1 is the graph of the zero map $S^2 \rightarrow S^2$ and E_2 is the graph of f ; then E_1 and E_2 have, homologically, a non-zero intersection number ($= n$), and therefore must actually intersect.

15. The “Fundamental Theorem of Algebra”

15.1. Theorem. *If $f(x)$ is a polynomial with complex coefficients, and if it has degree greater than 0, then there exists $a \in \mathcal{C}$ such that $f(a) = 0$.*

Several proofs: The first step in all the proofs given here is to make an estimate of the size of $|f(x)|$, to show that

$$\lim_{|x| \rightarrow \infty} |f(x)| = \infty.$$

Thus, we can compactify \mathcal{C} by adding ∞ , getting the topological space which is the 2-dimensional sphere S^2 . And we can extend f to a continuous function $f: S^2 \rightarrow S^2$ by defining $f(\infty) = \infty$.

1. By expanding f into a Taylor’s series around the point a , we get

$$f(x) = f(a) + b(x - a)^k + \text{terms of higher order in } (x - a).$$

We then use trigonometric formulas for the k th root of a complex number to show that f maps arbitrarily small neighborhoods of a onto a neighborhood of $f(a)$. This is also true near ∞ . Thus, $f(S^2)$ is an open subset of S^2 . On the other hand, S^2 is compact and Hausdorff, and thus $f(S^2)$ is a non-empty, closed subset of S^2 . Since S^2 is connected, this shows that $f(S^2) = S^2$, f is surjective, and hence takes on the value 0.

2. (Complex analysis) Since $f(S^2)$ is closed, if $0 \notin f(S^2)$, then f would not take on values close to 0, and hence $(f(x))^{-1}$ would be a bounded function, analytic in the entire complex plane. By Liouville’s Theorem, it would follow that $(f(x))^{-1}$ is constant; contradiction.

3. (Algebraic topology) The lower order coefficients of f can be joined by paths to 0, while the leading coefficient can be joined to 1 by a path not going through 0. This performs a homotopy of maps $S^2 \rightarrow S^2$, starting with f and ending with the map $x \rightarrow x^n$, where $n = \text{degree of } f$. Now, looking at the homology groups with coefficient group \mathbf{Z} , this proves that $f_*: H_2(S^2) \rightarrow H_2(S^2)$ is multiplication by n . But if f were not surjective, a simple construction shows that f would be null-homotopic, and so f_* would be multiplication by 0. We know that $H_2(S^2) = \mathbf{Z}$, and hence get a contradiction.

4. (Covering spaces) Let A be the set of complex numbers a for which the derivative $f'(a) = 0$, together with ∞ . This is a finite set, and, also, since f is at

most n to 1, we know that $f^{-1}(f(A))$ is finite. On $S^2 \setminus f^{-1}(f(A))$, the function f is a local homeomorphism. This, together with the fact that S^2 is compact, produces a covering projection $\tilde{f}: S^2 \setminus f^{-1}(f(A)) \rightarrow S^2 \setminus f(A)$. Since the target of \tilde{f} is path-connected, and the image of \tilde{f} is non-empty, we can, by lifting paths, show that \tilde{f} is surjective. ■

Proof (4) is the starting point of Smale's investigation into the computability of a zero of f using "Newton's method." It is the proof which is the most constructive. All these proofs are at heart the same. If mathematics is contractible, we naturally expect that any two proofs of the same thing are homotopic.

15.2. Corollary. *Every maximal ideal in $\mathcal{C}[x]$ is of the form $\langle x - a \rangle$.*

Another way to say this is to say that every non-constant polynomial $f(x)$ has a factor of the form $(x - a)$. Or, that the only irreducible complex polynomials are, up to constant factor, of that form. Certain other fields have this same property; they are said to be "algebraically closed."

16. Hilbert rings

- **16.1. Hilbert (or Jacobson) Rings.** A *Hilbert ring* is a comring in which every prime ideal is an intersection of maximal ideals. This is an important concept, which was invented by Goldman and Krull in their articles in *Math. Z.* (1951), in order to describe a generalization of a theorem of Hilbert's called the "Nullstellensatz." Goldman called them "Hilbert rings," while Krull called them "Jacobson rings." Kaplansky, in *Commutative rings*, calls them Hilbert rings; so does Bourbaki. But Atiyah and Macdonald in *Introduction to Commutative Algebra*, call them Jacobson rings.

- **16.2. R-domain.** I am now going to invent my own terminology, for temporary use: An *R-domain* is an integral domain in which the intersection of all maximal ideals is $\{0\}$. Thus, a comring A is a Hilbert ring, if and only if, for every prime ideal $P \subset A$, the quotient A/P is an R-domain. The "R" stands for (Jacobson) radical; the Jacobson radical of a (not necessarily commutative) ring A is the intersection of all maximal left-ideals of A . Jacobson found this concept particularly interesting, for non-commutative rings, and I think that is why Krull used the terminology "Jacobson ring."

- **16.3. Theorem (\mathbf{Z} and $F[x]$).** *The ring \mathbf{Z} is an R-domain. If F is a field, then $F[x]$ is an R-domain.*

Proof: If $|n| > 0$, then the number $|n| + 1$ has some prime factor p , and then $n \notin \langle p \rangle$. This idea is due to Euclid (*Elements*, Book IX, Proposition 20).

In $F[x]$, the non-zero constant polynomials are all in the complement of the maximal ideal $\langle x \rangle$. If $f(x) \in F[x]$ and has degree greater than 0, then $f(x) + 1$ has an irreducible factor $p(x)$, which generates a maximal ideal $\langle p(x) \rangle$ that does not contain $f(x)$. ■

A corollary is, since in these rings the only non-maximal prime ideal is $\{0\}$, that \mathbf{Z} and $F[x]$ are Hilbert rings.

16.4. $D[x]$. If D is an R-domain, then $D[x]$ is an R-domain.

Proof: Let $f(x) \in D[x] \setminus \{0\}$. The leading coefficient of $f(x)$ is non-zero, and hence is in the complement of some maximal ideal M in D (because D is an R-domain). Let F be the field D/M . Then f is non-zero in $F[x]$; hence, by **16.3**, there is a maximal ideal N of $F[x]$ which does not contain f . This maximal ideal pulls back to a maximal ideal of $D[x]$ that does not contain f . ■

Example: Consider those rational numbers which can be written in the form p/q , where p and q are in \mathbf{Z} , with $q \neq 0$ and with q not being divisible by 2 and not being divisible by 3. This forms a subring D of the field of rational numbers \mathbf{Q} . Now, D is a principal ideal domain, the proper ideals being generated by things of the form $2^k 3^n$. There are only two maximal ideals $\langle 2 \rangle$ and $\langle 3 \rangle$, whose intersection is $\langle 6 \rangle$. Thus, D is *not* an R-domain, and it is also not a Hilbert ring. This example can be generalized by taking any *finite* set of primes in place of $\{2, 3\}$; but if we took an *infinite* set of primes, we would get R-domains, Hilbert rings, as certain strange subdomains of \mathbf{Q} .

17. Integral extensions

• **17.1. Integral over a subring.** Let $A \subset B$ be comrings. An element $\beta \in B$ is said to be *integral over* A , if it satisfies a monic polynomial with coefficients in A . That is, if there is an integer $n \geq 1$ and there are elements a_0, a_1, \dots, a_{n-1} in A , such that

$$\beta^n + a_{n-1}\beta^{n-1} + \cdots + a_1\beta + a_0 = 0.$$

Note that every element of A is integral over A .

17.2. Theorem (Characterization of integral elements). *Let $A \subset B$ be comrings, and $\beta \in B$. Then β is integral over A , if and only if: There exists an A -submodule M of B , such that M is a finitely generated A -module, and $A \subset M$, and $\beta M \subset M$.*

Proof: If β satisfies a monic A -polynomial of degree n , then we can let M be the A -submodule of B generated by

$$\{1, \beta, \beta^2, \dots, \beta^{n-1}\}.$$

The point is that multiplication by β shifts these generators one to the right, and takes β^{n-1} to β^n , which is an A -linear combination of these generators, due to the fact that β^n can be solved for in the monic equation which β satisfies.

Conversely, let $1 = b_1, b_2, \dots, b_n$ generate M as an A -module. Then we can write

$$\beta b_i = \sum_{j=1}^n a_{ij} b_j$$

with $a_{ij} \in A$. Letting $X = (a_{ij})$ be the matrix of coefficients of these equations, and denoting by I the identity $n \times n$ matrix, and imagining (b_j) to be a column vector, we can write this

$$(\beta I - X) \cdot (b_j) = 0.$$

Multiply this on the left by the adjoint $(\beta I - X)^{\text{adj}}$, we get a diagonal matrix with $d = \det(\beta I - X)$ on the diagonal, multiplied times (b_j) , is equal to the column vector 0. Evaluate the first entry in the product, noting that $b_1 = 1$, and we see that $d = 0$. Now, $d = \det(\beta I - X)$ is just the evaluation of the “characteristic polynomial” of X at β ; and this polynomial is a monic polynomial with coefficients in A by **8.2**. ■

17.3. Corollary. *If $A \subset B$ are comrings, then the subset of B consisting of the elements which are integral over A is a subring of B containing A .*

Proof: Basically, what has to be shown is that if β and γ are two elements of B which are integral over A , then both $\beta + \gamma$ and $\beta\gamma$ are integral over A . Let $\{b_i\}$ be a finite set of elements of B , so that the A -module M generated by this set is mapped into itself when multiplying by β , and such that $b_1 = 1$. Let $\{c_j\}$ be similarly situated with reference to γ . Then it is easy to compute that the A -module N generated by $\{b_i c_j\}$ contains A and is mapped into itself by multiplication by $\beta + \gamma$ and by $\beta\gamma$. ■

We can therefore say that B is integral over A if every element of B is integral over A . And this can be verified if we only know that B is generated as a ring by A together with a set of elements each of which is integral over A .

17.4. Theorem. *Suppose that F is a field, and that F contains a subring D (which is, of course, an integral domain). And suppose that F is integral over D . Then D is a field.*

Proof: What must be shown is that if $d \in D \setminus \{0\}$, then the element $d^{-1} \in F$ in fact belongs to D . Write down what it means for d^{-1} to be integral over D : There are a_0, \dots, a_{n-1} in D , such that

$$(d^{-1})^n + a_{n-1}(d^{-1})^{n-1} + \cdots + a_1 d^{-1} + a_0 = 0.$$

Multiply by d^{n-1} and solve for d^{-1} ; you will get an expression for d^{-1} as an element of D . ■

18. Localization

Let D be an integral domain, containing an element $c \neq 0$. We consider fractions whose numerators belong to D and whose denominators are non-negative powers of c . That is, we consider pairs a/c^n , for $a \in D$, $n \geq 0$. We define an equivalence relation on such pairs, so that a/c^k and b/c^n are equivalent if and only if $a \cdot c^n = b \cdot c^k$ is a true equation in D . We then define addition and multiplication of equivalence classes in the usual way. What results is an integral domain, denoted by D_c ; this contains D , and it is a subring of the field of fractions of D .

* EXERCISE 18.1: Another way to describe D_c is as $D[x]/\langle cx - 1 \rangle$; this description needs a little lemma, that $\langle cx - 1 \rangle$ is a prime ideal in $D[x]$.

18.1. Theorem. *Let D be an integral domain, $c \in D \setminus \{0\}$. Let A be a comring, and let $\varphi: D \rightarrow A$ be a ring-map. Suppose that $\varphi(c)$ is a unit in A , with inverse a . Then φ extends in a unique way to a ring-map $: D_c \rightarrow A$; the extension takes the fraction d/c^n to $\varphi(d) \cdot a^n$. ■*

* EXERCISE 18.2: The proof is a simple calculation.

18.2. Corollary (Localizing a maximal ideal). *Let D be an integral domain, M a maximal ideal in D , and $c \in D \setminus M$. Then $M_c = D_c \cdot M$, defined in an obvious way, is a maximal ideal in D_c ; and $D_c/M_c \approx D/M$; furthermore, if $M_c = \{0\}$, then $M = \{0\}$. ■*

* EXERCISE 18.3: Go through the details.

18.3. Theorem (Localizing an R-domain). *Let D be an R-domain, and suppose that $c \in D \setminus \{0\}$. Then D_c is an R-domain.*

Proof: Consider a non-zero element d/c^n of D_c . The element $dc \in D$ is non-zero. Therefore, there is a maximal ideal M of D , which does not contain the element dc (because D is an R-domain). It follows that neither c nor d belongs to M ; and hence M_c is a maximal ideal of D_c which does not contain the element d/c^n . ■

18.4. Theorem. *Suppose that D is an R-domain, and that there is a maximal ideal M in $D[x]$, such that $M \cap D = \{0\}$. Then D is a field.*

Proof: (We obviously mean, by saying that D is a subring of $D[x]$, to embed D as the set of polynomials of degree 0.) First of all, the maximal ideal M cannot be $\{0\}$, since, for instance, the polynomial x has no multiplicative inverse in $D[x]$. So, let $f(x)$ be a non-zero element of M ; this is a polynomial of positive degree, and has a leading coefficient which we call c . Localize at c . We get $D_c \subset D_c[x]$, and M_c is a maximal ideal in $D_c[x]$. Furthermore, let β be the image of x in the quotient $F = D_c[x]/M_c$, a field. Then $c^{-1}f(\beta) = 0$ exhibits a monic polynomial, with coefficients in D_c , which β satisfies. Since F is generated as a ring by D_c and β , it follows from **17.3** that F is integral over D_c ; and hence, from **17.4**, we have the interesting fact that D_c is a field.

Now, let N be a maximal ideal in D , such that $c \notin N$. This exists because D is an R-domain. Then, by **18.2**, N_c is a maximal ideal in D_c ; but this latter is a field and has only the maximal ideal $\{0\}$; thus, $N_c = \{0\}$, and hence by **18.2**, $N = \{0\}$. And so D is a field, since $\{0\}$ is a maximal ideal in it. ■

19. The Goldman-Krull Theorem, Part 1

19.1. Theorem. *Let A be a Hilbert ring. Let M be a maximal ideal in $A[x]$. Then $A \cap M$ is a maximal ideal in A .*

Proof: Define $D = A/(A \cap M)$; this is a subring of the field $A[x]/M$, and is therefore an integral domain; being a quotient domain of a Hilbert ring, it is an R-domain. Now, there is an image of M in the polynomial domain $D[x]$; let us call it M' . We have $D \cap M' = \{0\}$, and $D[x]/M'$ is isomorphic to $A[x]/M$, a field; thus, M' is a maximal ideal in $D[x]$. It follows from **18.4** that D is a field, and so $A \cap M$ is a maximal ideal in A . ■

20. Extending maximal ideals in integral extensions

• **20.1. Algebraic elements.** Suppose that $D \subset E$ are integral domains. An element $\beta \in E$ is said to be *algebraic* over D , when there is a non-zero polynomial $f(x) \in D[x]$ such that $f(\beta) = 0$. By localizing at the leading coefficient of f , we can arrange various results relating this notion to the notion of “integral.” An element of E which is not algebraic over D is called *transcendental*. We call E algebraic over D when every element of E is algebraic over D .

20.2. Theorem. *Suppose that $D \subset E$ are integral domains with E algebraic over D . If I is any ideal of E such that $D \cap I = \{0\}$, then $I = \{0\}$.*

Proof: Suppose that $\beta \in I$ and $\beta \neq 0$. Since β is algebraic over D , it satisfies a polynomial in $D[x]$ of minimal degree:

$$a_n\beta^n + \cdots + a_1\beta + a_0 = 0,$$

where $a_i \in D$. The claim is that $a_0 \neq 0$. Otherwise, because E is a domain and $\beta \neq 0$, we could cancel a factor β and get a polynomial of smaller degree. But then, we can solve the above equation for a_0 , and find that $a_0 \in D \cap I$. ■

20.3. Theorem (Max ideal extends into integral extension). *Let $A \subset B$ be comrings, with B integral over A . Suppose that M is a maximal ideal in A . Then there is a maximal ideal \mathcal{M} of B , such that $\mathcal{M} \cap A = M$.*

Proof: This uses a clever little trick. We consider the set of those ideals N of B such that $N \cap A \subset M$. Of such ideals, there is a maximal one \mathcal{M} , by Zorn’s Lemma. The claim is that $\mathcal{M} \cap A = M$. Admit that for a minute; then any larger ideal in B must intersect A in something larger than M , and hence must intersect A in all of A , hence must contain 1, hence must be all of B ; and so \mathcal{M} will be a maximal ideal in B .

Thus, the problem is to show that if N is a B -ideal, with $N \cap A \subset M$, and if there exists an element $\mu \in M \setminus N$, then the B -ideal N' generated by N together with μ , has the property that $N' \cap A \subset M$. Let $\alpha = \nu + b\mu$ be an arbitrary element of $N' \cap A$, where $\nu \in N$ and $b \in B$. Since b is integral over A , there is an equation, with $a_i \in A$:

$$b^n = a_{n-1}b^{n-1} + \cdots + a_1b + a_0.$$

Multiply this equation by μ^n , and be aware that $b\mu$ equals α modulo N :

$$(b\mu)^n = \mu \cdot (a_{n-1}(b\mu)^{n-1} + \cdots + a_1(b\mu)\mu^{n-2} + a_0\mu^{n-1}).$$

$$\alpha^n = \mu \cdot (a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha\mu^{n-2} + a_0\mu^{n-1}) + \nu_1,$$

where $\nu_1 \in N$. Now, look at this carefully: $\nu_1 \in N$ and $\nu_1 \in A$. From the assumption that $N \cap A \subset M$, it follows that $\nu_1 \in M$. The rest of the expression on the right of the equation is a multiple of μ by an element of A ; and since $\mu \in M$, it follows that the whole right side of the equation belongs to M . Thus, α^n is zero modulo M ; but, since A/M is a field, this implies that α is zero modulo M ; that is, that $\alpha \in M$. ■

20.4. Corollary. *If $D \subset E$ are integral domains, with E generated by D together with one element β which is algebraic over D , and if D is an R-domain, then E is an R-domain.*

Proof: Localize at the leading coefficient c of a D -polynomial which β satisfies. Then D_c is an R-domain, by **18.3**, and E_c is integral over D_c . Let I be the intersection of all maximal ideals of E_c . Since each maximal ideal of D_c is the intersection of a maximal ideal of E_c with D_c , by **20.3**, we can conclude that $I \cap D_c$ is contained in the intersection of all maximal ideals of D_c , which is $\{0\}$ because D_c is an R-domain. And then **20.2** implies that $I = \{0\}$. The proof is finished off by noting that the intersection of all maximal ideals of E is contained in I . ■

21. The Goldman-Krull Theorem, Part 2

21.1. Theorem. *If A is a Hilbert ring, then $A[x]$ is a Hilbert ring.*

Proof: Let P be a prime ideal in $A[x]$. Define $D = A/(A \cap P)$. This is contained in $D[x]$, which maps to $A[x]/P$ with kernel P' . That is, $D[x]/P' \approx A[x]/P$. Now, D is an R-domain contained in $E = D[x]/P'$.

If $P' = \{0\}$, then, by **16.4**, E is an R-domain.

Otherwise, E is generated by D and the element β which is the image of x modulo P' . Since we are now assuming that $P' \neq \{0\}$, we see that β is algebraic over D . We conclude from **20.4** that E is an R-domain.

In other words, in all cases, E is an R-domain; the intersection of all maximal ideals in $E \approx A[x]/P$ is zero. Hence the intersection of all maximal ideals in $A[x]$ which contain P is exactly P . ■

22. Hilbert's Nullstellensatz

22.1. Theorem. *Every maximal ideal in $\mathcal{C}[x_1, \dots, x_n]$ is of the form*

$$\langle x_1 - a_1, \dots, x_n - a_n \rangle,$$

for some choice of complex numbers a_1, \dots, a_n .

Proof: The proof is by induction on n . For $n = 1$, this is the Fundamental Theorem of Algebra **15.2**.

Assume this theorem for $n - 1$. By part 2 of the Goldman-Krull Theorem **21.1**, we know that $\Lambda = \mathcal{C}[x_1, \dots, x_{n-1}]$ is a Hilbert ring. Let M be a maximal ideal in $\Lambda[x_n]$. We know, by part 1 of the Goldman-Krull Theorem **19.1**, that $\Lambda \cap M$ is a maximal ideal in Λ . Hence, by the inductive assumption, $\Lambda \cap M$ is generated by $\{x_1 - a_1, \dots, x_{n-1} - a_{n-1}\}$; furthermore, from this it follows that $\mathcal{C} \rightarrow \Lambda/(\Lambda \cap M)$ is an isomorphism. Factor $\Lambda[x_n]$ by its ideal generated by $\Lambda \cap M$, and we get $\mathcal{C}[x_n]$ in which the image of M is a maximal ideal M' . By the Fundamental Theorem of Algebra, M' is generated by $\{x_n - a_n\}$. We can now work back to a set of generators of M as an ideal in $\Lambda[x_n]$ and reach the desired conclusion. ■

• **22.2. Variety of an ideal.** Given an ideal $I \subset \mathcal{C}[x_1, \dots, x_n]$, we define the *variety of I* to be a subset of \mathcal{C}^n :

$$V(I) = \{(z_1, \dots, z_n) \in \mathcal{C}^n \mid \forall p \in I, p(z_1, \dots, z_n) = 0\}.$$

Note that if $I \subset J$, then $V(J) \subset V(I)$.

22.3. Corollary. *If M is a maximal ideal in $\mathcal{C}[x_1, \dots, x_n]$, then $V(M)$ consists of one point $\{(a_1, \dots, a_n)\}$ in \mathcal{C}^n .*

22.4. Corollary. *If I is an ideal in $\mathcal{C}[x_1, \dots, x_n]$ and $V(I) = \emptyset$, then $1 \in I$.*

Proof: If I were a proper ideal (i.e., $1 \notin I$), then, by Zorn's Lemma, there would be a maximal ideal M containing I . Then $\emptyset \neq V(M) \subset V(I)$. ■

22.5. Theorem [Hilbert's form of the "Nullstellensatz"]. *Suppose that I is an ideal in $\mathcal{C}[x_1, \dots, x_n]$, and let $g \in \mathcal{C}[x_1, \dots, x_n]$, such that, for all $\mathbf{z} \in V(I)$, we have $g(\mathbf{z}) = 0$. Then there is some positive integer N such that $g^N \in I$.*

Proof: By the Hilbert Basis Theorem **4.1**, some finite set $\{f_1(\mathbf{x}), \dots, f_k(\mathbf{x})\}$ generates I . Adjoin an extra symbol y , getting $\Lambda = \mathcal{C}[x_1, \dots, x_n, y]$. Consider the ideal

J in Λ generated by $\{f_1(\mathbf{x}), \dots, f_k(\mathbf{x}), 1 - y \cdot g(\mathbf{x})\}$. Whenever $\{\mathbf{x}, y\}$ is zeroed by f_1, \dots, f_n , then $1 - y \cdot g(\mathbf{x})$ has the value 1. Thus $V(J) = \emptyset$. By **22.4**, we can write:

$$1 = \left(\sum_{i=1}^k h_i(\mathbf{x}, y) f_i(\mathbf{x}) \right) + r(\mathbf{x}, y) \cdot (1 - y \cdot g(\mathbf{x}))$$

where $h_i, r \in \Lambda$. Now, make the substitution $y = (g(\mathbf{x}))^{-1}$, getting an equation valid in the field of fractions of $\mathcal{C}[x_1, \dots, x_n]$, in which the denominators are all powers of g . Multiplying by a large power of g we get:

$$(g(\mathbf{x}))^N = \sum_{i=1}^k \left(g^N \cdot h_i \left(\mathbf{x}, \frac{1}{g(\mathbf{x})} \right) \right) \cdot f_i(\mathbf{x})$$

where the coefficients $g^N \cdot h_i(\mathbf{x}, g^{-1})$ belong to $\mathcal{C}[x_1, \dots, x_n]$. ■

This argument remains valid when \mathcal{C} is replaced by any algebraically closed field. One can say something about arbitrary fields as well. For example, the proof of **22.1** can be modified to show that, if F is any field, then every maximal ideal in $F[x_1, \dots, x_n]$ can be generated by n elements.

23. Finitely generated rings

The aim of this section is to show how to use the Goldman-Krull theorem to prove the result (10.6) proved earlier in a more primitive manner, that a finitely generated comring is residually finite.

23.1. Theorem. *If M is any maximal ideal in $\Lambda = \mathbf{Z}[x_1, \dots, x_n]$, then Λ/M is a finite field.*

Proof: By the Goldman-Krull theorem, part 2 **21.1**, starting with the fact that \mathbf{Z} is a Hilbert ring, we know that Λ is a Hilbert ring. We now work backwards, using Goldman-Krull, part 1 **19.1**, and get to the point where $\mathbf{Z} \cap M$ is a maximal ideal in \mathbf{Z} . Hence, for some prime number p , $\mathbf{Z} \cap M = \langle p \rangle$. Continuing forward again, suppose we call $\mathbf{Z}[x_1, \dots, x_k] = \Lambda_k$; we inductively show that $F_k = \Lambda_k / (M \cap \Lambda_k)$ is a finite field. To get $\Lambda_{k+1} / (M \cap \Lambda_{k+1})$, we factor $F_k[x_k]$ by a maximal ideal; this gives F_{k+1} as a finite-dimensional vector space over F_k (—N.B., $\{0\}$ is *not* a maximal ideal in $F_k[x_{k+1}]$ —), and so F_{k+1} is finite. ■

By a *finitely generated comring*, one means a ring A isomorphic to $\mathbf{Z}[x_1, \dots, x_n]$ modulo some ideal. The Hilbert Basis Theorem **4.1** implies that $\mathbf{Z}[x_1, \dots, x_n]$ is Noetherian, and so its quotient A is also Noetherian.

23.2. Theorem (Finitely generated comrings are residually finite). *If A is any finitely generated comring, and if $a \in A \setminus \{0\}$, then there is an ideal J of A such that $a \in A \setminus J$ and A/J is a finite ring.*

Proof: By **9.3**, a corollary of the Krull Intersection Theorem, there is a maximal ideal M of A and an integer $n \geq 1$, such that $a \in A \setminus M^n$. The claim is that A/M^n is finite. For $n = 1$, this is implied by **23.1**, since the inverse image of M in $\mathbf{Z}[x_1, \dots, x_k]$ is of “finite index.” Now, M^n is an ideal in the Noetherian ring A , hence generated by finitely many elements; the quotient M^n/M^{n+1} has the structure of an (A/M) -module; it is thus a finite-dimensional vector space over the finite field A/M , hence a finite set. The number of elements $|X|$ of X satisfies

$$\left| \frac{A}{M^{n+1}} \right| = \left| \frac{A}{M^n} \right| \cdot \left| \frac{M^n}{M^{n+1}} \right|. \blacksquare$$

• **23.3. Residually finite group.** We call a group G “residually finite” when, for each $g \in G \setminus \{1\}$, there is a normal subgroup N of G such that $g \in G \setminus N$ and G/N is finite. Clearly, every subgroup of a residually finite group is residually finite.

Given some positive integer n , we define $GL_n(A)$ to be the group of invertible $n \times n$ matrices over the comring A .

23.4. Theorem. (a) *If A is a finitely generated comring, then $GL_n(A)$ is residually finite.*

(b) *If G is a finitely generated subgroup of $GL_n(B)$, where B is a comring, then G is residually finite.*

Proof: (a) Suppose $\alpha \in GL_n(A)$ and $\alpha \neq I$; then the matrix α differs from the identity matrix I at some entry. By **23.2**, A has an ideal J , such that A/J is finite and such that, modulo J , it is still true that $\alpha \neq I$. We can use the quotient homomorphism $\varphi: A \rightarrow A/J$ to define a homomorphism $\tilde{\varphi}: GL_n(A) \rightarrow GL_n(A/J)$ in which $\tilde{\varphi}(\alpha) \neq I$. Now, $GL_n(A/J)$ is a finite group; hence $N = \ker \tilde{\varphi}$ is a normal subgroup of finite index in $GL_n(A)$ with $\alpha \notin N$.

(b) Suppose G is generated by the matrices $\{g_1, \dots, g_k\}$. Let A be the subring of B generated by all the entries in the matrices

$$\{g_1, \dots, g_k, g_1^{-1}, \dots, g_k^{-1}\}.$$

Then A is a finitely generated comring. The homomorphism

$$GL_n(A) \rightarrow GL_n(B)$$

is injective; its image contains G . Hence G is isomorphic to a subgroup of $GL_n(A)$, which is residually finite by part (a). ■