Lin Xun

Commutative Algebra and Basic Algebraic Geometry

Undergraduate Lecture Notes

MATH3005

The Chinese University of Hong Kong, Shenzhen

https://linlinsai.github.io/index.html

Version 7 · October 2026

Lin Xun

The Chinese University of Hong Kong, Shenzhen
Shenzhen, China

Homepage:
https://linlinsai.github.io/index.html

MATH3005
Version 7
October 2026

Preface

Commutative algebra studies rings and the equations expressed in them. Its elementary examples come from arithmetic and from polynomials. Algebraic geometry begins with the solution sets of polynomial equations and studies the maps between them. The purpose of these notes is to develop the algebra carefully and then explain how it describes geometric objects.

The notes assume familiarity with elementary algebra, complex numbers, and linear algebra. No algebraic geometry is assumed. Lecture 1 begins with symmetries of planar figures, extracts the group axioms from their compositions, and reviews groups, rings, and polynomials. New definitions are placed before the propositions and calculations that use them. Examples explain the definitions, and proofs identify the hypotheses needed at each step.

The first lecture draws on the supplied *Geometry and Symmetry* notes [3]. Section 1.6 develops conjugation and Cauchy's theorem as applications of group actions. Lectures 2–12 develop the commutative algebra of Atiyah–Macdonald [1]. The subsequent lectures follow the geometric topics in the MATH3005 syllabus [4], using Hartshorne [2] as the principal geometric reference.

Primary decomposition and completion are outside the scope of these notes. Appendix B lists the 155 included numbered main-text results from Atiyah–Macdonald and gives the location of each statement and proof. It also lists the selected source exercises. Their statements are restated in our own words. Each lecture ends with exercises, and their proofs and solutions appear in the separate companion volume, *Exercise Solutions*. Results needed in the main development are proved in the exposition. Lecture 24 includes a complete descent proof of Mordell's theorem over $\mathbb{Q}$, with the required arithmetic lemmas. Some further geometric theorems are stated without full proofs; these are identified where they occur.

A useful way to study the notes is to work out each example independently and then compare the calculation with the text. For each map, write down its source and target. For each theorem, record its hypotheses before following the proof. In particular, check why a proposed operation on equivalence classes does not depend on the representatives. Try the exercises at the end of each lecture before consulting the companion solutions.

Exercise numbers agree in both volumes; for example, Exercise 8.3 belongs to Lecture 8. The examples and theorem proofs remain part of the lecture text.

This revised edition expands the explanations within all 26 lectures. Additional calculations verify well-definedness, exhibit the role of hypotheses, and connect the algebraic constructions with local geometric examples. Section titles name their mathematical subjects directly. The original exercise numbers and the source-result labels are retained.

There are 26 lectures, arranged as two units for each teaching week of the supplied syllabus. The table records this correspondence; the syllabus itself specifies weekly topics rather than individual meetings. Only weeks containing mathematical lectures are listed. The mathematical development follows this order, with elementary preliminary definitions supplied whenever a later topic is needed in an earlier argument.

The syllabus and the lectures

Week	Lectures	Topics	Main reading
1	1–2	Abstract algebra review; rings, ideals, homomorphisms, zero divisors, prime and maximal ideals, radicals, Chinese remainder theorem.	Li §1.1–1.5; AM Ch. 1
2	3–4	Modules, generators and relations, free modules, submodules, quotients, direct sums and products, exact sequences.	AM Ch. 2
3	5–6	Tensor products, exactness, algebras, restriction and extension of scalars.	AM Ch. 2
4	7–8	Rings and modules of fractions, local properties, local rings, Nakayama's lemma.	AM Chs. 2–3
5	9–10	Chain conditions, Noetherian and Artinian rings, integral dependence, valuation rings and discrete valuations.	AM Chs. 5–9 (selected scope)
6	11–12	Graded rings and modules, Hilbert functions, dimension, local Noetherian rings.	AM Chs. 10–11 (no completion)
7	13–14	Categories and abelian categories, complexes, cohomology, long exact sequences.	AM Ch. 2; these notes
9	15–16	Basic topology, manifolds, categories of spaces; Zariski topology and prime spectra.	Hartshorne I.1; II.2
10	17–18	Nullstellensatz, affine varieties, morphisms, rational maps, relative viewpoint; first look at projective coordinates and blow-ups.	Hartshorne I.1, I.3–4
11	19–20	Projective and quasi-projective varieties, intersections, Hilbert polynomials and plane Bézout.	Hartshorne I.2, I.7
12	21–22	Algebraic curves, normalization, local calculations, and detailed blow-up charts.	Hartshorne I.5–6; II.7
13	23–24	Elliptic curves, the group law, rational points, and the complete elementary proof of Mordell's theorem.	Hartshorne IV.4; Milne IV
14	25–26	Sheaves, quasi-coherent and coherent sheaves, vector bundles, projective examples; optional cohomology on $\mathbb{P}^1$.	Hartshorne II.1, II.5; III.4–5

The blow-up is previewed in the syllabus's Week 10 position and computed fully after curves in Week 12. This keeps its first appearance on schedule while allowing the later calculation to use normalization, local rings, and projective charts.

Contents

Preface iii
The syllabus and the lectures v

1 Review: symmetry, maps, quotients, and polynomial rings 1
1.1 Symmetry and the definition of a group 1
1.2 Group calculations and elementary arithmetic 5
1.3 Group homomorphisms 9
1.4 Cosets and Lagrange's theorem 10
1.5 Quotient groups and the first isomorphism theorem 11
1.6 Group actions 13
1.7 Rings and fields 16
1.8 Polynomials and quotient rings 18
1.9 A quotient ring associated to a parabola 20
1.10 Exercises 22

2 Ideals, prime ideals, and the arithmetic of equations 23
2.1 Ideals and quotient rings 23
2.2 Prime and maximal ideals 25
2.3 Nilpotent elements and radicals 28
2.4 Ideal operations and the Chinese remainder theorem 30
2.5 Avoidance, quotients, and radical identities 31
2.6 Extension and contraction of ideals 35
2.7 Arithmetic of polynomial coefficients 37
2.8 Boolean rings and idempotents 38
2.9 The prime spectrum 39
2.10 Exercises 40

3 Modules: linear algebra over a ring 42
3.1 Modules and module homomorphisms 42
3.2 Submodules, quotients, and the first isomorphism theorem 43
3.3 Generators, bases, and presentations 44
3.4 Polynomial modules and linear operators 47

3.5 Computing maps of modules 48
3.6 Isomorphism theorems and annihilators 49
3.7 Exercises 51

4 Direct sums and exact sequences 52
4.1 Direct sums and direct products 52
4.2 Exact sequences 53
4.3 Split exact sequences 54
4.4 Exactness after taking Hom 56
4.5 Exactness criteria using Hom 57
4.6 The snake lemma 58
4.7 Additivity and finite generation in exact sequences 60
4.8 Exercises 60

5 Tensor products: making bilinear maps linear 62
5.1 Bilinear maps and the tensor product 62
5.2 Construction by generators and relations 63
5.3 Induced maps and basic tensor identities 64
5.4 Tensoring a quotient 65
5.5 Right exactness of tensor products 66
5.6 Flat modules and failure of left exactness 67
5.7 Finite witnesses and several tensor factors 68
5.8 The finite test for flatness 70
5.9 Exercises 71

6 Algebras and change of scalars 72
6.1 Algebras and their structure maps 72
6.2 Restriction of scalars 73
6.3 Extension of scalars 73
6.4 Changing scalars in a presentation 75
6.5 Tensor products of algebras and equations 75
6.6 Fiber algebras 76
6.7 Finiteness under changing scalars 78
6.8 Tensor products over two rings 79
6.9 Exercises 80

7 Localization: choosing which elements may be denominators 81
7.1 Multiplicative subsets and rings of fractions 81
7.2 The universal property 83
7.3 Modules and exactness 84
7.4 Prime ideals and local tests 85
7.5 Recognizing and combining localizations 87
7.6 Localization of module operations 87
7.7 Local criteria 89
7.8 Extension and contraction under localization 89
7.9 Exercises 91

8 Local rings and Nakayama's lemma 93
8.1 Local rings and residue fields 93
8.2 Reduction modulo the maximal ideal 94
8.3 Minimal generators 97
8.4 Consequences and hypotheses of Nakayama's lemma 98
8.5 Locality and finite generation 98
8.6 Support of a module 99
8.7 Exercises 100

9 Noetherian and Artinian rings 102
9.1 Chain conditions 102
9.2 Polynomial rings and Hilbert's basis theorem 105
9.3 Finite algebras and the weak Nullstellensatz 107
9.4 Composition series and length 111
9.5 Nilpotent ideals and minimal primes 114
9.6 Artinian rings and their local factors 115
9.7 Finiteness of spectra and fibers 119
9.8 Exercises 120

10 Valuations, integral dependence, and normalization 122
10.1 Order of vanishing as arithmetic 122
10.2 Monic equations and finite modules 124
10.3 Polynomial factorization and integral closure 127
10.4 The integral closure of the cusp ring 129
10.5 Lying over and going up 129
10.6 Localization, normality, and going down 131
10.7 Finite separable extensions and the trace 133
10.8 Existence of valuation rings 137
10.9 Base change and finite projections 139
10.10 Recognizing a discrete valuation ring 141
10.11 Dedekind domains and ideal factorization 143
10.12 Fractional ideals and their inverses 145
10.13 Exercises 149

11 Graded rings and Hilbert functions 151
11.1 Graded rings and modules 151
11.2 Hilbert functions and generating series 152
11.3 The general Hilbert–Serre theorem 154
11.4 Filtrations and associated graded rings 157
11.5 Stable filtrations and the Artin–Rees lemma 158
11.6 Exercises 162

12 Dimension theory and regular local rings 163
12.1 Prime chains and height 163
12.2 The Hilbert–Samuel polynomial 164
12.3 Growth under quotient by a non-zero-divisor 166
12.4 Systems of parameters 167
12.5 Height bounds and the effect of one equation 169
12.6 Independence of parameters 170
12.7 Regular local rings and tangent spaces 172
12.8 Transcendence degree and the dimension of a variety 175
12.9 Worked Hilbert–Samuel calculations 178
12.10 Exercises 179

13 Categories, functors, and complexes 181
13.1 Categories and functors 181
13.2 Natural transformations and universal properties 182
13.3 Complexes and cohomology 184
13.4 Exact sequences of complexes 186
13.5 Exercises 187

14 The long exact sequence in cohomology 188
14.1 The connecting homomorphism 188
14.2 A two-term complex over the integers 190
14.3 Reduction modulo an integer: a second connecting map 190
14.4 Naturality of the connecting homomorphism 191
14.5 Exercises 192

15 A first course in topology and manifolds 193
15.1 Topological spaces and continuous maps 193
15.2 Closure, density, connectedness, and separation 194
15.3 Covers, compactness, and products 195
15.4 Manifolds and coordinate charts 196
15.5 Exercises 198

16 The Zariski topology and the spectrum of a ring 199
16.1 Algebraic sets and the Zariski topology 199
16.2 Irreducible spaces and components 201
16.3 The prime spectrum 202
16.4 Exercises 205

17 The Nullstellensatz and affine algebraic sets 206
17.1 Vanishing ideals and zero sets 206
17.2 Zariski's lemma 207
17.3 Weak and strong forms of the Nullstellensatz 208
17.4 Coordinate rings and irreducibility 210
17.5 Exercises 211

18 Regular maps, rational maps, and families 212
18.1 Regular functions and local rings 212
18.2 Morphisms and pullbacks 214
18.3 Rational maps and birational equivalence 215
18.4 Families and fiber rings 217
18.5 Projective directions and an incidence variety 218
18.6 Exercises 219

19 Projective varieties and the meaning of intersection 220
19.1 Projective coordinates 220
19.2 Equations and affine charts 221
19.3 Homogenization and projective closure 222
19.4 Maps to projective space 224
19.5 Local intersection multiplicity 225
19.6 Exercises 226

20 Hilbert polynomials and Bézout's theorem 227
20.1 Hilbert functions and degree 227
20.2 The Hilbert polynomial of a hypersurface 228
20.3 The Hilbert series of a plane intersection 229
20.4 The Hilbert polynomial and total local length 229
20.5 Examples of Bézout's theorem 231
20.6 Arithmetic genus 232
20.7 Exercises 233

21 Algebraic curves, singularities, and normalization 234
21.1 Curves and their function fields 234
21.2 Tangent space and the local criterion for singularity 235
21.3 Integral closure and normalization 236
21.4 Normalization of the cusp 237
21.5 Normalization of the node 238
21.6 Arithmetic genus and geometric genus 240
21.7 Exercises 241

22 Blowing up a point: complete calculations in two charts 243
22.1 The incidence construction 243
22.2 The two affine charts and their transition 244
22.3 Total and strict transforms 245
22.4 Blowing up the cusp 245
22.5 Blowing up the node 247
22.6 The Rees algebra and relative projective construction 248
22.7 Exercises 250

23 Elliptic curves and the chord–tangent group law 251
23.1 The equation and its point at infinity 251
23.2 Secant and tangent formulas 252
23.3 Vertical lines, the identity, and torsion 254
23.4 Divisors and associativity 256
23.5 A local calculation at infinity 258
23.6 Exercises 259

24 Rational points, heights, and Mordell's theorem 260
24.1 Rational point calculations 260
24.2 Finitely generated groups and Mordell's theorem 262
24.3 Heights and descent 264
24.4 The arithmetic needed for descent 264
24.5 Weak Mordell: reducing to finitely many square classes 270
24.6 Height estimates and the proof of Mordell's theorem 275
24.7 Counting points over a finite field 281
24.8 Exercises 281

25 Sheaves: local information and gluing 283
25.1 Presheaves and sheaves 283
25.2 Germs and stalks 284
25.3 The structure sheaf of a spectrum 285
25.4 Skyscraper sheaves and morphisms 287
25.5 Kernels, cokernels, and exactness 288
25.6 Exercises 289

26 Modules become sheaves; vector bundles on the projective line 290
26.1 The sheaf associated to a module 290
26.2 Quasi-coherent and coherent sheaves 291
26.3 Locally free sheaves and their fibers 293
26.4 Constructing $\mathscr{O}(n)$ on $\mathbb{P}^1$ 294
26.5 Čech cohomology on the projective line 296
26.6 Exercises 299

A Notation and a compact algebra–geometry dictionary 301

B Atiyah–Macdonald: source-result guide 303

References 311

Lecture 1
Review: symmetry, maps, quotients, and polynomial rings

A rectangular window can be reflected, a triangular road sign can be rotated, and a square tile can be turned over. In each case the shape remains unchanged, but its points move. We begin by calculating with these movements; the group axioms will record the rules that the calculations reveal. The examples follow the viewpoint of [3, Sections 1.1–1.5]. Later in the chapter, polynomial substitution leads to quotient rings and the equation of a parabola.

1.1 Symmetry and the definition of a group

Place a rectangular window in the coordinate plane. Reflection u in its horizontal midline exchanges its top and bottom; reflection v in its vertical midline exchanges its left and right. Performing both reflections produces a half-turn.

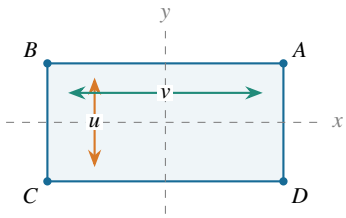


Fig. 1.1 The reflections u and v of a rectangle. Their composite is a half-turn.

Here $\mathbb{R}$ denotes the real numbers, and $\mathbb{R}^2$ is the set of ordered pairs (x,y) of real numbers. The notation $P \in X$ means that P is an element of X; $H \subseteq X$ means that every element of H belongs to X. Braces list the elements of a set. A *map* $f : X \to Y$ assigns one point $f(P) \in Y$ to each point $P \in X$. A point P is a *preimage* of Q if $f(P) = Q$. The map is a *bijection* if every point of Y has exactly one preimage. Its inverse f^{-1} sends $f(P)$ back to P. For maps $f : X \to Y$ and $g : Y \to Z$, define *composition* by

$$(g\circ f)(P) = g(f(P)).$$

The rightmost map acts first. We also write $gf = g\circ f$. The *identity map* e is given by $e(P) = P$.

The distance between $P = (x,y)$ and $Q = (x',y')$ is

$$d(P,Q) = \sqrt{(x-x')^2 + (y-y')^2}.$$

A *symmetry* of a plane figure X is a bijection $f : X \to X$ satisfying

$$d(f(P), f(Q)) = d(P,Q) \qquad (P,Q \in X).$$

Thus a symmetry preserves both the figure and all distances within it.

Example 1.1 (Two reflections of a rectangle)**.** For the rectangle $-2 \le x \le 2$, $-1 \le y \le 1$, put

$$\begin{aligned} e(x,y) &= (x,y), & u(x,y) &= (x,-y),\\ v(x,y) &= (-x,y), & w(x,y) &= (-x,-y). \end{aligned}$$

Each map preserves the rectangle and distances. For example,

$$d(u(P),u(Q))^2 = (x-x')^2 + (-y+y')^2 = d(P,Q)^2.$$

Direct substitution gives

$$uu = vv = ww = e, \qquad uv = vu = w, \quad uw = wu = v, \quad vw = wv = u.$$

All composites therefore belong to the same set $\{e,u,v,w\}$. The complete multiplication table is

$\circ$	e	u	v	w
e	e	u	v	w
u	u	e	w	v
v	v	w	e	u
w	w	v	u	e

The entry in row f, column g is $fg = f\circ g$. Every movement can be undone: in this example its inverse is the same movement.

Example 1.2 (Rotating a triangular sign)**.** Let X be the equilateral triangle with vertices

$$A = (1,0), \qquad B = \left(-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\right), \qquad C = \left(-\tfrac{1}{2}, -\tfrac{\sqrt{3}}{2}\right).$$

Counterclockwise rotation through 120° is the map

$$\rho(x,y) = \left(-\tfrac{1}{2}x - \tfrac{\sqrt{3}}{2}y, \tfrac{\sqrt{3}}{2}x - \tfrac{1}{2}y\right).$$

It sends $A \mapsto B \mapsto C \mapsto A$. Write $\rho^2 = \rho\circ\rho$ and $\rho^3 = \rho\circ\rho\circ\rho$. Three turns restore every point:

$$\rho^3 = e, \qquad \rho\rho^2 = \rho^2\rho = e, \qquad \rho^2\rho^2 = \rho.$$

Thus $\{e, \rho, \rho^2\}$ is closed under composition, with

$$e^{-1} = e, \qquad \rho^{-1} = \rho^2, \qquad (\rho^2)^{-1} = \rho.$$

A symmetry need not be its own inverse. These are the three rotational symmetries; the triangle also has three reflections.

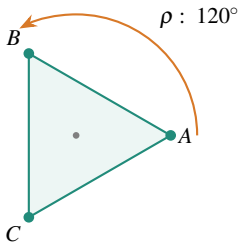


Fig. 1.2 Three rotations of 120° return every point to its starting position.

Example 1.3 (A square tile: the order of the movements matters)**.** For the square with vertices

$$1 = (1,1), \quad 2 = (-1,1), \quad 3 = (-1,-1), \quad 4 = (1,-1),$$

let r be counterclockwise rotation through 90°, and let s be reflection in the horizontal axis:

$$r(x,y) = (-y,x), \qquad s(x,y) = (x,-y).$$

Here r^j means j successive rotations, with $r^0 = e$. Then

$$\begin{aligned} r^4 &= e, & s^2 &= e, \\ sr(x,y) &= (-y,-x), & rs(x,y) &= (y,x). \end{aligned}$$

For instance,

$$\begin{aligned} (1,0) &\xrightarrow{r} (0,1) \xrightarrow{s} (0,-1), \\ (1,0) &\xrightarrow{s} (1,0) \xrightarrow{r} (0,1). \end{aligned}$$

Consequently $sr \neq rs$: doing the same two movements in a different order can change the result. More precisely,

$$sr = r^3 s.$$

Every composite of r and s can be reduced to one of

$$e, r, r^2, r^3, \qquad s, rs, r^2 s, r^3 s.$$

Indeed, move each s to the right using $sr = r^3 s$, cancel each pair $ss = e$, and reduce the number of rotations using $r^4 = e$. These eight maps are distinct: the first four preserve the

cyclic order of the vertices, the last four reverse it, and within either list the image of vertex 1 distinguishes the maps.

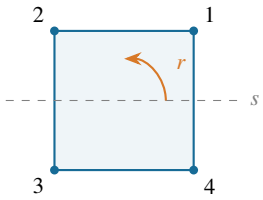


Fig. 1.3 r sends $1 \mapsto 2 \mapsto 3 \mapsto 4 \mapsto 1$; s exchanges $1 \leftrightarrow 4$ and $2 \leftrightarrow 3$.

1.1.1 The common rules

Let G be the set of all symmetries of a figure X. For $f, g, h \in G$ and $P, Q \in X$, composition satisfies four rules.

(1) *Closure.* The unique preimage of Q under fg is $g^{-1}(f^{-1}(Q))$, so fg is a bijection. It preserves distances:

$$d(fg(P), fg(Q)) = d(g(P), g(Q)) = d(P, Q), \qquad fg \in G.$$

(2) *Associativity.* Both parenthesizations have the same value:

$$((fg)h)(P) = f(g(h(P))) = (f(gh))(P).$$

(3) *Identity.* Doing nothing has no effect:

$$e(P) = P, \qquad ef = fe = f.$$

(4) *Inverse.* Undoing a symmetry is a symmetry:

$$f^{-1}f = ff^{-1} = e, \qquad d(f^{-1}(P), f^{-1}(Q)) = d(P, Q).$$

The distance identity follows by applying the distance-preserving map f to the two points on its left.

The square shows that $fg = gf$ is not a necessary rule. We retain exactly the four rules above and allow the elements to be objects other than movements of a figure. In the definition, $G \times G$ denotes the set of ordered pairs of elements of G.

Definition 1.4. A *group* is a set G with an operation

$$G \times G \longrightarrow G, \qquad (g,h) \longmapsto gh,$$

satisfying the following axioms:

$$\begin{aligned} (gh)k = g(hk) \quad & \text{for all } g,h,k \in G; \\ eg = ge = g \quad & \text{for some } e \in G \text{ and all } g \in G; \\ gg^{-1} = g^{-1}g = e \quad & \text{for some } g^{-1} \in G \text{ for each } g \in G. \end{aligned}$$

The element e is the *identity*, and g^{-1} is an *inverse* of g. Closure is included in the requirement that the operation have target G. The group is *abelian* if

$$gh = hg \qquad (g,h \in G).$$

A *subgroup* is a nonempty subset $H \subseteq G$ such that

$$h,k \in H \quad \Longrightarrow \quad hk \in H, \quad h^{-1} \in H.$$

It contains $e = hh^{-1}$ and inherits a group structure; $\{e\}$ is the *trivial subgroup*. For groups G,H, a *group isomorphism* is a bijection $f : G \to H$ satisfying

$$f(gg') = f(g)f(g') \qquad (g,g' \in G).$$

If such a map exists, we write $G \simeq H$ and call the groups *isomorphic*.

The rectangle gives an abelian group; the three rotations of the triangle give another. The square gives a nonabelian group, denoted D_4 here, following [3]; some books denote the same eight-element group by D_8.

1.2 Group calculations and elementary arithmetic

1.2.1 Notation, maps, and elementary arithmetic

We use $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C}$ for the integers, rationals, reals, and complex numbers. A complex number has the form $a+bi$, where $a,b \in \mathbb{R}$ and $i^2 = -1$. The basic set notation is

$$\varnothing = \text{the empty set}, \qquad |X| = \text{the number of elements of a finite set } X.$$

$$\begin{aligned} X \cup Y &= \{x : x \in X \text{ or } x \in Y\}, \\ X \cap Y &= \{x : x \in X \text{ and } x \in Y\}, \\ X \setminus Y &= \{x : x \in X,\ x \notin Y\}, \\ X \times Y &= \{(x,y) : x \in X,\ y \in Y\}. \end{aligned}$$

The inclusion $X \subseteq Y$ means that every element of X belongs to Y; $X \subsetneq Y$ also requires $X \neq Y$. The set X^n consists of ordered lists of n elements of X. For any indexed family of sets,

$$x \in \bigcup_\lambda X_\lambda \iff x \in X_\lambda \text{ for some } \lambda,$$
$$x \in \bigcap_\lambda X_\lambda \iff x \in X_\lambda \text{ for every } \lambda.$$

Sums and products abbreviate repeated operations:

$$\sum_{i=1}^n a_i = a_1 + \cdots + a_n, \qquad \prod_{i=1}^n a_i = a_1 \cdots a_n.$$

For a map $f : X \to Y$, the sets X, Y are its *source* and *target*. The following conditions distinguish three useful types of map:

$$\begin{aligned} f \text{ injective} &\iff f(x) = f(x') \text{ implies } x = x'; \\ f \text{ surjective} &\iff \text{each } y \in Y \text{ equals } f(x) \text{ for some } x \in X; \\ f \text{ bijective} &\iff f \text{ is injective and surjective.} \end{aligned}$$

For subsets $E \subseteq X$ and $F \subseteq Y$, its *image* and *inverse image* are

$$f(E) = \{f(x) : x \in E\}, \qquad f^{-1}(F) = \{x \in X : f(x) \in F\}.$$

The set $f^{-1}(\{y\})$ is the *fiber* over y. Inverse images are defined even when an inverse function does not exist. We also write id_X for the identity map on X.

Example 1.5. For $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$,

$$f(1) = f(-1), \qquad f(\mathbb{R}) = [0,\infty), \qquad f^{-1}(\{4\}) = \{-2,2\}, \qquad f^{-1}(\{-1\}) = \varnothing.$$

Thus f is neither injective nor surjective. Restricting its source and target gives the bijection

$$[0,\infty) \longrightarrow [0,\infty), \quad x \longmapsto x^2, \qquad f^{-1}(y) = \sqrt{y}.$$

A relation $\sim$ on X is an *equivalence relation* if it satisfies

$$\begin{aligned} x \sim x & \qquad \text{(reflexivity)}, \\ x \sim y \implies y \sim x & \qquad \text{(symmetry)}, \\ x \sim y,\ y \sim z \implies x \sim z & \qquad \text{(transitivity)}. \end{aligned}$$

The *equivalence class* of x is $[x] = \{y \in X : y \sim x\}$. These classes partition X:

$$x \in [x], \qquad [x] \cap [y] \neq \varnothing \implies [x] = [y].$$

For the second assertion, a common member gives $x \sim y$ by symmetry and transitivity; then $z \sim x$ is equivalent to $z \sim y$. A formula on classes is *well defined* if its value is independent of the representative. For example, the proposed rule $[x] \mapsto F(x)$ requires

$$x \sim x' \quad \Longrightarrow \quad F(x) = F(x').$$

We use the well-ordering property of the nonnegative integers: every nonempty subset has a least element. Its equivalent induction principle proves a statement for every nonnegative integer by proving the case 0 and showing that the statement for n implies it for $n+1$. These principles justify the least-element and induction arguments below.

For integers a, b, we write $a \mid b$ if $b = ac$ for some integer c. A positive integer $p > 1$ is *prime* if its only positive divisors are $1, p$. The greatest common divisor $\gcd(a,b)$ of two integers not both zero is their largest positive common divisor. They are *relatively prime* if this divisor is 1.

Proposition 1.6 (Integer division and Bézout's identity)**.** *For $a \in \mathbb{Z}$ and $d > 0$, there are unique integers q, r such that $a = qd + r$ and $0 \le r < d$. If a, b are not both zero, there are integers u, v with*

$$\gcd(a,b) = ua + vb.$$

Consequently, if $\gcd(a,b) = 1$ *and $a \mid bc$, then $a \mid c$.*

Proof. Choose the smallest nonnegative member r of the set $\{a - qd : q \in \mathbb{Z},\ a - qd \ge 0\}$. This set is nonempty. If $r \ge d$, then $r - d$ is a smaller member, a contradiction. For uniqueness, the difference of two possible remainders is a multiple of d of absolute value less than d, hence is zero.

For Bézout's identity, choose the least positive integer d expressible as $ua + vb$; such a positive integer exists by changing the sign of a nonzero one of a, b. Divide a by d. Its remainder is again an integer combination of a, b, so minimality forces that remainder to be zero. Thus $d \mid a$, and similarly $d \mid b$. Every common divisor of a, b divides their combination d, so $d = \gcd(a,b)$. Finally, multiplying $ua + vb = 1$ by c shows that $a \mid c$ whenever $a \mid bc$. □

For example, $2 = 2 \cdot 8 - 14$ shows $\gcd(8,14) = 2$. If p is prime and $p \nmid a$, the final assertion with $\gcd(p,a) = 1$ shows that $p \mid ab$ forces $p \mid b$.

Every integer $n > 1$ is a product of primes: if n is not prime, factor it as ab with $1 < a, b < n$ and apply induction to the factors. This factorization is unique apart from the order of the primes. Indeed, a prime in one factorization must divide a prime in the other by the preceding divisibility assertion, hence equals that prime; cancel and repeat. This is the *fundamental theorem of arithmetic*. In particular, every positive divisor of a prime power p^a is itself a power of p.

Proposition 1.7 (The elementary rules behind group calculations)**.** *The identity and the inverse of each element of a group are unique. Cancellation holds on either side. Moreover*

$$(ab)^{-1} = b^{-1}a^{-1}, \qquad (a^{-1})^{-1} = a.$$

Proof. If e and e' are both identities, then $e = ee' = e'$. If b and c are both inverses of a, then

$$b = be = b(ac) = (ba)c = ec = c.$$

If $ab = ac$, multiply on the left by a^{-1} to obtain $b = c$; the proof for right cancellation is the same with the factors on the other side. Finally

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = e,$$

and the product in the other order is also e. Uniqueness of the inverse proves the first formula. Since a is an inverse of a^{-1}, it proves the second formula too. □

The parentheses in this proof are justified by associativity. None of the steps interchanges two unrelated factors, so the proof does not assume commutativity. The distinction between reassociating factors and interchanging them will matter throughout the group calculations.

An isomorphism $f : G \to H$ sends identity to identity: $f(e) = f(e)^2$, so cancellation in H gives $f(e) = e$. Applying f to $gg^{-1} = e$ then gives $f(g^{-1}) = f(g)^{-1}$. These facts justify comparing identities, inverses, and powers in isomorphic groups.

For an integer $m > 0$, the power g^m means a product of m copies of g. Set $g^0 = e$ and $g^{-m} = (g^{-1})^m$. Associativity gives $g^{a+b} = g^a g^b$ for integers a, b. The *order* of an element g is the least positive integer n such that $g^n = e$, if one exists; otherwise its order is infinite. The *order* of a finite group is its cardinality.

The integers form an abelian group under addition: the identity is 0 and the inverse of a is $-a$. The nonzero rational numbers form a group under multiplication. All rational numbers do not form a multiplicative group, because 0 has no inverse.

The square symmetries r, s from Section 1.1 also have a matrix description. A matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ represents the map $(x, y) \mapsto (ax + by, cx + dy)$. A matrix is *invertible* if its map has an inverse represented by a matrix; equivalently, it has a two-sided matrix inverse. Matrix multiplication represents composition, and $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ represents the identity. In this notation,

$$r = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Multiply these matrices to check

$$r^4 = I, \qquad s^2 = I, \qquad sr = r^{-1}s.$$

Every product can therefore be rewritten as r^i or $r^i s$, with $0 \le i < 4$. These are the eight transformations of D_4: four rotations and four reflections.

Example 1.8 (The order of multiplication matters)**.** Apply sr to $(1, 0)$. Rotation gives $(0, 1)$, and reflection gives $(0, -1)$. Applying rs gives $(0, 1)$. Hence $sr \neq rs$. In a product of transformations, the rightmost transformation acts first. Matrix groups are useful examples even though most rings later in this course will be commutative.

1.2.2 Cyclic groups and direct products

The subgroup

$$\langle g \rangle = \{g^m : m \in \mathbb{Z}\}$$

is called the cyclic subgroup generated by g. A group is *cyclic* if it equals $\langle g\rangle$ for some g. If g has order n, integer division reduces every exponent modulo n, and $e, g, \dots, g^{n-1}$ are distinct. Thus $\langle g\rangle$ has n elements. In the square group, r has order four, while every reflection has order two.

Example 1.9 (Clock arithmetic and roots of unity)**.** For an integer $n \geq 1$, define $a \equiv b \pmod{n}$ to mean $n \mid a-b$. Divisibility shows that this is an equivalence relation on $\mathbb{Z}$. Write $\bar{a} = a + n\mathbb{Z}$ for its class and $\mathbb{Z}/n\mathbb{Z}$ for the set of classes. Integer division gives the representatives $\bar{0}, \bar{1}, \dots, \overline{n-1}$. Define addition by $\bar{a} + \bar{b} = \overline{a+b}$. This is well defined: changing either representative by a multiple of n changes the sum by a multiple of n. Addition is associative, $\bar{0}$ is its identity, and $-\bar{a} = \overline{-a}$ is the inverse. The resulting additive group is cyclic, generated by $\bar{1}$. The complex numbers

$$\mu_n = \{z \in \mathbb{C} : z^n = 1\} = \{1, \zeta, \zeta^2, \dots, \zeta^{n-1}\}, \qquad \zeta = \cos(2\pi/n) + i\sin(2\pi/n),$$

form a cyclic multiplicative group. Multiplication by ζ rotates the complex plane through $2\pi/n$. The map $\bar{a} \mapsto \zeta^a$ identifies clock addition with composition of rotations. It is well defined because adding n to the exponent does not change the value.

Example 1.10 (Two groups with four elements)**.** For groups G, H, the *direct product* $G \times H$ uses coordinatewise multiplication: $(g,h)(g',h') = (gg', hh')$. For additive groups, the rule is coordinatewise addition. The group $\mathbb{Z}/4\mathbb{Z}$ has an element of order four. The group

$$V = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$$

has no such element: adding any element to itself gives zero. Write its elements as $0, a, b, c$, where $a = (\bar{1}, \bar{0})$, $b = (\bar{0}, \bar{1})$, and $c = a + b$. Then

$$a + a = b + b = c + c = 0, \qquad a + b = c, \quad b + c = a, \quad c + a = b.$$

This is the Klein four group. Equal cardinalities do not imply isomorphic groups: for an isomorphism f, the equality $f(g)^m = f(g^m)$ and injectivity show that $f(g)^m = e$ if and only if $g^m = e$. Thus element orders are preserved.

1.3 Group homomorphisms

A *group homomorphism* is a map $\varphi : G \to H$ satisfying $\varphi(gg') = \varphi(g)\varphi(g')$. It sends the identity to the identity: $\varphi(e) = \varphi(e)^2$, so cancellation gives $\varphi(e) = e$. It also sends inverses to inverses, by applying it to $gg^{-1} = e$.

The *kernel* is $\ker\varphi = \{g : \varphi(g) = e\}$, and the *image* is $\operatorname{im}\varphi = \{\varphi(g) : g \in G\}$. The kernel is closed under products and inverses because φ preserves them. The image is closed under these operations for the same reason. Thus both are subgroups. A homomorphism is injective precisely when its kernel is $\{e\}$. Indeed, if $\varphi(g) = \varphi(h)$, then $\varphi(h^{-1}g) = e$. A trivial kernel forces $g = h$; the converse follows immediately from injectivity. A bijective homomorphism is exactly a group isomorphism. Its inverse also preserves multiplication: if $y = \varphi(g)$ and $z = \varphi(h)$, then $\varphi^{-1}(yz) = gh = \varphi^{-1}(y)\varphi^{-1}(z)$.

Example 1.11 (An integer forgets all but its remainder)**.** For an integer $n \geq 2$, the map $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ sends a to its congruence class $\bar{a}$. The equality $\overline{a+b} = \bar{a} + \bar{b}$ makes it a homomorphism of additive groups. It is surjective and has kernel $n\mathbb{Z}$. Its fibers consist of numbers differing by a multiple of n.

Example 1.12 (A linear-algebra example)**.** Let $\mathrm{GL}_2(\mathbb{R})$ denote the group of invertible 2×2 real matrices. For $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ its determinant is $\det M = ad - bc$. If $ad - bc \neq 0$, direct multiplication gives the inverse

$$M^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Expanding the four entries of a product gives $\det(MN) = \det M \det N$. This identity also shows that an invertible matrix has nonzero determinant. Hence the determinant gives a homomorphism

$$\det : \mathrm{GL}_2(\mathbb{R}) \longrightarrow \mathbb{R}^\times,$$

where $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$ with multiplication. It is surjective because $\det \begin{pmatrix} u & 0 \\ 0 & 1 \end{pmatrix} = u$. Its kernel is the subgroup $\mathrm{SL}_2(\mathbb{R})$ of matrices of determinant one. The identity in this target group is 1, so belonging to the kernel means having determinant 1.

1.4 Cosets and Lagrange's theorem

We now apply the equivalence-class construction to a subgroup. Unlike the congruence classes of integers, these classes need not inherit a group operation.

Let $H \subseteq G$ be a subgroup. A *left coset* is $gH = \{gh : h \in H\}$. If aH and bH meet, write $ah_1 = bh_2$. Then $b^{-1}a = h_2 h_1^{-1} \in H$, and multiplying shows $aH = bH$. Thus distinct left cosets are disjoint. Each has the same number of elements as H, since $h \mapsto gh$ is a bijection, with inverse $gh \mapsto h$. The set of left cosets is written G/H, and its cardinality $[G : H] = |G/H|$ is called the *index* of H in G.

Theorem 1.13 (Lagrange)**.** *If G is finite, then $|G| = |G/H|\,|H|$, where G/H here means the set of left cosets. In particular $|H|$ divides $|G|$.*

Proof. Every $g \in G$ belongs to gH. The preceding argument partitions G into $|G/H|$ disjoint sets, each of cardinality $|H|$. Count their elements. □

This is [3, Theorem 1.20]. It does not assert that every divisor of $|G|$ occurs as a subgroup order.

Corollary 1.14 (Orders of elements)**.** *If G is finite and $g \in G$, then the order of g divides $|G|$, and $g^{|G|} = e$. A group of prime order is cyclic.*

Proof. Apply Lagrange's theorem to $\langle g\rangle$. If $|G| = p$ is prime and $g \neq e$, the subgroup $\langle g\rangle$ has more than one element and its order divides p. It must have p elements, so it equals G. □

Theorem 1.15 (Subgroups and classification of cyclic groups)**.** *Every subgroup of a cyclic group is cyclic. Every infinite cyclic group is isomorphic to $(\mathbb{Z},+)$, and every cyclic group of order n is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.*

Proof. Let $G = \langle g\rangle$ and let H be a subgroup. If $H = \{e\}$, it is generated by e. Otherwise some nonidentity power of g lies in H. By replacing a negative exponent by its negative and taking the inverse, we obtain a positive exponent. Choose the least positive integer d such that $g^d \in H$.

For any $g^m \in H$, integer division gives $m = qd + r$ with $0 \leq r < d$. Since H is closed under products and inverses,

$$g^r = g^m (g^d)^{-q} \in H.$$

Minimality of d implies $r = 0$. Thus every element of H is a power of g^d, and $H = \langle g^d\rangle$.

For classification, the map $\mathbb{Z} \to G$, $m \mapsto g^m$, is onto. If g has infinite order, equality $g^m = g^r$ forces $m = r$, so this is an isomorphism. If g has order n, equality holds exactly when n divides $m - r$. Consequently $\bar{m} \mapsto g^m$ is a well-defined bijective homomorphism $\mathbb{Z}/n\mathbb{Z} \to G$. □

This is the result of [3, Proposition 1.26], together with the explicit description of cyclic groups. The proof uses the same "choose the smallest positive element, then divide" idea that will classify ideals in $\mathbb{Z}$ and in a polynomial ring in one variable.

1.5 Quotient groups and the first isomorphism theorem

Cosets always form a set. To make that set into a group, we must check that multiplication does not depend on the chosen representatives. To multiply cosets, we would like $(gH)(g'H) = gg'H$. For $g, h \in G$, the element ghg^{-1} is called the *conjugate* of h by g, and gHg^{-1} is the set of these conjugates with $h \in H$. A subgroup H is *normal* if $gHg^{-1} = H$ for every g. In this case the proposed multiplication works. To see the role of normality, replace g, g' by $gh, g'h'$ with $h, h' \in H$. Then

$$(gh)(g'h') = gg'(g'^{-1}hg')h' \in gg'H.$$

The bracketed factor lies in H exactly because conjugation preserves H. The cosets then form the *quotient group* G/H: associativity follows from that of G, the identity is $H = eH$, and $(gH)^{-1} = g^{-1}H$. If G is abelian, then $ghg^{-1} = h$; hence every subgroup of an abelian group is normal. Every kernel is normal, since $\varphi(ghg^{-1}) = \varphi(g)e\varphi(g)^{-1} = e$ for $h \in \ker\varphi$.

Theorem 1.16 (First isomorphism theorem for groups)**.** *For a group homomorphism $\varphi : G \to H$, there is an isomorphism*

$$G/\ker\varphi \longrightarrow \operatorname{im}\varphi, \qquad g\ker\varphi \longmapsto \varphi(g).$$

Proof. Write $K = \ker\varphi$. If $gK = g'K$, then $g'^{-1}g \in K$, so $\varphi(g) = \varphi(g')$. Thus the displayed rule is well defined. It preserves multiplication by the homomorphism property of φ. Every element of the image has a preimage. Finally, equality of two output values implies $g'^{-1}g \in K$ and hence equality of the two cosets. This proves injectivity as well as surjectivity. □

The theorem says that a map becomes injective once we identify precisely the elements it cannot distinguish. The same construction will later apply to other algebraic objects.

Why normality is exactly the condition needed for a quotient

The quotient construction has two separate tasks: describing its elements and defining multiplication on those elements. Left cosets exist for every subgroup $H \subseteq G$, but the proposed formula

$$(gH)(hH) = (gh)H$$

does not always define an operation. To check it, replace g by ga and h by hb, with $a, b \in H$. Their product is

$$(ga)(hb) = gh\,(h^{-1}ah)b.$$

The new representative belongs to ghH provided $h^{-1}ah \in H$. Normality guarantees this for every $h \in G$ and $a \in H$. Conversely, if the formula defines a group operation with the canonical projection $\pi : G \to G/H$ a homomorphism, then $H = \ker\pi$ and

$$\pi(gag^{-1}) = \pi(g)\pi(a)\pi(g)^{-1} = 1 \quad (a \in H).$$

Thus H must be normal. This proves necessity as well as sufficiency.

For a concrete failure, use $G = S_3$ and $H = \{1, (12)\}$. Set $h = (123)$ and use the rightmost-permutation-first convention. The representatives 1 and (12) describe the same element H, but multiplying them by h gives the cosets hH and $(12)hH$. These are unequal: equality would require

$$h^{-1}(12)h = (13) \in H,$$

which is false. The difficulty is not associativity; multiplication has already failed to be well defined.

The first isomorphism theorem can now be read as an explicit identification. If $\varphi : G \to K$ is a homomorphism, put $N = \ker\varphi$ and define

$$\overline{\varphi} : G/N \longrightarrow \operatorname{im}\varphi, \qquad gN \longmapsto \varphi(g).$$

If $gN = g'N$, then $g' = gn$ for $n \in N$, so the formula is independent of the representative. If $\overline{\varphi}(gN) = 1$, then $g \in N$, so its kernel is trivial. Every element of $\operatorname{im}\varphi$ has the required form, so it is onto. Each conclusion comes from a different part of the construction.

1.6 Group actions

An action of G on a set X assigns a point gx to (g,x) with $ex = x$ and $(gh)x = g(hx)$. Each g acts by a bijection, whose inverse is the action of g^{-1}. The *orbit* of x is $Gx = \{gx : g \in G\}$; the *stabilizer* is $G_x = \{g : gx = x\}$.

Proposition 1.17 (Stabilizers at different points)**.** *For an action of G on X, each stabilizer G_x is a subgroup, and*

$$G_{gx} = gG_xg^{-1}.$$

The orbits partition X into disjoint subsets.

Proof. The identity fixes x. If h and k fix x, then $hkx = hx = x$; if h fixes x, applying h^{-1} gives $h^{-1}x = x$. This proves the subgroup property. For the conjugation formula,

$$h \in G_{gx} \iff hgx = gx \iff g^{-1}hgx = x \iff g^{-1}hg \in G_x.$$

Finally, if $y = gx$, every point hy equals $(hg)x$, so $Gy \subseteq Gx$. Applying g^{-1} gives the reverse inclusion. Hence intersecting orbits are equal, and every point lies in its own orbit. □

Proposition 1.18 (Orbit–stabilizer)**.** *The map $G/G_x \to Gx$, $gG_x \mapsto gx$, is a bijection. If G is finite, then $|Gx| = |G|/|G_x|$.*

Proof. If $g' = gh$ with $h \in G_x$, then $g'x = g(hx) = gx$, proving well-definedness. Conversely $g'x = gx$ implies $g^{-1}g' \in G_x$, so the cosets are equal. Surjectivity is the definition of orbit. The counting formula follows from Lagrange's theorem. □

For the square, all four vertices lie in one orbit. A given vertex is fixed by the identity and reflection in its diagonal. Hence the stabilizer has order two and the orbit has order $8/2 = 4$. This also illustrates why a stabilizer need not be normal: adjacent vertices have distinct stabilizers, conjugated into one another by r, so these stabilizers are not normal.

These are the three parts of [3, Theorem 1.58], with the subgroup and orbit-partition checks included. An action is *transitive* if it has just one orbit. It is *free* if every stabilizer is trivial, and *faithful* if the only group element fixing every point is the identity. Freeness asks about each individual stabilizer; faithfulness asks about their intersection. The square action on its four vertices is faithful but not free.

A *permutation* of a set X is a bijection $X \to X$. Such bijections form a group under composition, denoted S_X; for $X = \{1,\dots,n\}$ we write S_n. Choosing the image of the first symbol in n ways, the next in $n-1$ ways, and so on shows that $|S_n| = n! = n(n-1)\cdots 1$. An action is equivalently a homomorphism $G \to S_X$: the action law says that the permutation for gh is the composite of the permutations for g and h.

For calculations on a finite set, *cycle notation* records the successive images of each symbol. The cycle $(1\ 3\ 2)$ sends 1 to 3, 3 to 2, and 2 back to 1, fixing all other symbols. Thus $(1\ 3\ 2)(4\ 5) \in S_5$ makes a three-step circuit and swaps 4 with 5. A two-element cycle is called a *transposition*. As with matrices, the rightmost permutation acts first.

Every permutation of a finite set decomposes into disjoint cycles. To find them, start with a symbol and repeatedly apply the permutation. Because the set is finite and the map is bijective, the first repeated symbol is the starting one. This gives one cycle. Start again

with an unused symbol until every symbol is included. Disjoint cycles commute, since they move different symbols. Fixed symbols count as cycles of length one, though they are usually omitted from the notation. The square rotation, for example, is $(1\ 2\ 3\ 4)$, and its horizontal reflection is $(1\ 4)(2\ 3)$.

Theorem 1.19 (Cayley's theorem)**.** *Every group G is isomorphic to a subgroup of the permutation group of its underlying set. If $|G| = n$, it is isomorphic to a subgroup of S_n.*

Proof. For $g \in G$, let $L_g : G \to G$ be left multiplication, $L_g(h) = gh$. It is a bijection with inverse $L_{g^{-1}}$. Associativity gives

$$L_g L_{g'}(h) = g(g'h) = (gg')h = L_{gg'}(h),$$

so $g \mapsto L_g$ is a homomorphism $G \to S_G$. If $L_g = L_{g'}$, evaluating at e gives $g = g'$. Thus the map is injective. Its image is a subgroup isomorphic to G by the first isomorphism theorem. □

This is the faithful left-translation action described in [3, Example 1.47]. It shows that an abstract group can always be realized as symmetries of a set. It does not say that every group of order n is the full group S_n, whose order is usually much larger.

Example 1.20 (Write down the permutation representation)**.** For the Klein group $V = \{0, a, b, c\}$, left addition by a acts as $(0\ a)(b\ c)$; left addition by b acts as $(0\ b)(a\ c)$; and left addition by c acts as $(0\ c)(a\ b)$. Thus Cayley's theorem gives a concrete subgroup of S_4 with four elements. Every nonidentity element is a product of two disjoint transpositions.

Remark 1.21 (Preview: points of projective space). The multiplicative group $\mathbb{R}^\times$ acts on the set of nonzero real vectors $\mathbb{R}^{n+1} \setminus \{0\}$ by $a(x_0, \ldots, x_n) = (ax_0, \ldots, ax_n)$. Its orbits identify vectors differing by a common nonzero scalar. Write an orbit as $[x_0 : \cdots : x_n]$; the set of these orbits is called real *projective space*, denoted $\mathbb{P}^n(\mathbb{R})$. For example, $[1 : 2] = [2 : 4]$, while $[1 : 2] \neq [1 : 3]$. Each orbit is one point of this new set. Lecture 19 will develop its geometry and extend the construction to other coefficient fields. No quotient-group structure is claimed for this orbit set.

1.6.1 Conjugation and the class equation

Besides acting by left multiplication, a group acts on itself by *conjugation*:

$$g \cdot h = ghg^{-1}.$$

The action law follows from $g_1(g_2 h g_2^{-1})g_1^{-1} = (g_1 g_2)h(g_1 g_2)^{-1}$. The orbit of h is its *conjugacy class*, and its stabilizer is the *centralizer*

$$C_G(h) = \{g \in G : gh = hg\}.$$

The *center* is $Z(G) = \{h \in G : hg = gh \text{ for every } g \in G\}$. The center is a normal subgroup: products and inverses of elements commuting with everything still commute with everything, and conjugation fixes each central element.

Theorem 1.22 (Conjugacy of permutations)**.** *Two permutations in S_n are conjugate if and only if their disjoint-cycle decompositions have the same list of cycle lengths, counted with multiplicity.*

Proof. For a cycle, evaluate on each symbol to obtain

$$\tau(i_1\ i_2\ \cdots\ i_r)\tau^{-1} = (\tau(i_1)\ \tau(i_2)\ \cdots\ \tau(i_r)).$$

Conjugation therefore relabels each cycle and preserves its length. Conversely, suppose two permutations have matching cycle lengths. Match their cycles of each length and match the entries in their cyclic order. These matches define a bijection τ of all n symbols. The displayed formula shows that conjugating the first permutation by τ gives the second. □

This is [3, Theorem 1.56]. For example, the conjugacy classes of S_3 are

$$\{e\}, \qquad \{(12),(13),(23)\}, \qquad \{(123),(132)\}.$$

Their sizes are 1, 3, and 2. The centralizer of (12) has size $6/3 = 2$, while the centralizer of (123) has size $6/2 = 3$.

Theorem 1.23 (The class equation)**.** *For a finite group G, choose one representative x_i from each conjugacy class with more than one element. Then*

$$|G| = |Z(G)| + \sum_i [G : C_G(x_i)],$$

where $[G : H] = |G/H|$ is the index *of a subgroup.*

Proof. Conjugacy classes partition G. The one-element classes are exactly the central elements, because $\{ghg^{-1} : g \in G\} = \{h\}$ means that h commutes with every g. Orbit–stabilizer gives the size $[G : C_G(x_i)]$ for each remaining class. Summing the sizes proves the equation. □

Corollary 1.24 (Finite groups of prime-power order)**.** *If $|G| = p^a$ for a prime p and $a \geq 1$, then p divides $|Z(G)|$, so the center is nontrivial. Every group of order p^2 is abelian.*

Proof. Each noncentral class has size greater than one and dividing p^a, so its size is divisible by p. Reducing the class equation modulo p gives $|Z(G)| \equiv 0 \pmod{p}$. The center contains e and therefore has at least p elements.

For $|G| = p^2$, Lagrange leaves only $|Z(G)| = p$ or p^2. In the latter case G is abelian. In the former, $G/Z(G)$ has prime order and is cyclic. Choose g whose coset generates this quotient. Every element of G then has the form $g^r z$ with $z \in Z(G)$. Two such elements commute, since

$$(g^r z)(g^s w) = g^{r+s} zw = g^{s+r} wz = (g^s w)(g^r z).$$

Thus G is again abelian. In fact this contradicts $|Z(G)| = p$, so only the second case occurs. □

The class equation is useful because the numerical sizes of orbits force algebraic structure. It turns a counting argument into a theorem about commutativity.

1.6.2 Cauchy's theorem by rotating a list

Theorem 1.25 (Cauchy)**.** *If a prime p divides the order of a finite group G, then G contains an element of order p.*

Proof. Consider the finite set

$$T = \{(g_1, \dots, g_p) \in G^p : g_1 g_2 \cdots g_p = e\}.$$

The first $p-1$ entries can be chosen arbitrarily, and the last must be $(g_1 \cdots g_{p-1})^{-1}$. Therefore $|T| = |G|^{p-1}$, which is divisible by p.

Rotate the entries one place:

$$(g_1, g_2, \dots, g_p) \longmapsto (g_2, \dots, g_p, g_1).$$

This really stays inside T, even if G is nonabelian, because

$$g_2 \cdots g_p g_1 = g_1^{-1}(g_1 \cdots g_p) g_1 = e.$$

After p rotations the list returns to itself. Thus the cyclic group of order p acts on T. Its orbits have size either one or p, by orbit–stabilizer. The one-element orbits are precisely the constant lists $(g, \dots, g)$ with $g^p = e$. Write F for the number of these lists. Counting the remaining orbits gives $|T| = F + pm$ for an integer m, so F is divisible by p.

The constant list $(e, \dots, e)$ exists, so $F \geq 1$. Divisibility forces $F \geq p$. There is therefore some $g \neq e$ with $g^p = e$. Its order divides p and is not one, hence is exactly p. □

This result is a partial converse to Lagrange: every *prime* divisor of the group order occurs as an element order. The adjective prime is essential. For instance the Klein four group has order four but no element of order four.

1.7 Rings and fields

Definition 1.26. A *commutative ring* A is a set with addition and multiplication such that $(A, +)$ is an abelian group, multiplication is associative and commutative with an identity 1, and

$$a(b+c) = ab + ac, \qquad (a+b)c = ac + bc.$$

The additive identity is denoted 0. A *unit* is an element u such that $uv = 1$ for some $v \in A$. A *field* is a commutative ring with $0 \neq 1$ in which every nonzero element is a unit. An *integral domain* is a commutative ring with $0 \neq 1$ in which $ab = 0$ implies $a = 0$ or $b = 0$.

The identity $a0 = a(0+0) = a0 + a0$ gives $a0 = 0$ by additive cancellation. If $0 = 1$, then every a equals $a1 = a0 = 0$; this one-element ring is the *zero ring*. We allow it as a ring but exclude it from fields and domains. In any other ring, 0 cannot be a unit. Units form an abelian group under multiplication: if $uv = 1$, the inverse v is unique by the associativity calculation used to prove uniqueness of group inverses, and a product of units is a unit. This group is denoted $A^\times$.

A *zero divisor* is an element $a \in A$ for which $ab = 0$ for some nonzero $b \in A$. With this convention 0 is a zero divisor in every nonzero ring. A domain has no nonzero zero divisors. An element is *nilpotent* if $a^m = 0$ for some positive integer m. These two conditions differ: a nonzero nilpotent is a zero divisor, but a zero divisor need not have a zero power.

A *ring homomorphism* $f : A \to B$ satisfies

$$f(a+a') = f(a)+f(a'), \quad f(aa') = f(a)f(a'), \quad f(1_A) = 1_B.$$

It also sends 0 to 0 and negatives to negatives, by the additive-group identities. Its kernel is $\ker f = \{a : f(a) = 0\}$, and its image is $f(A)$. A bijective ring homomorphism is a *ring isomorphism*; its inverse preserves the operations by the same argument as for groups. A *subring* of B is a subset containing 1_B and closed under addition, negatives, and multiplication, with the inherited operations.

The integers form a domain, and $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. The units of $\mathbb{Z}$ are $1, -1$. A field is a domain because $ab = 0$ and $a \neq 0$ give $b = a^{-1}ab = 0$. The converse fails: 2 has no inverse in $\mathbb{Z}$. In a domain one may cancel a nonzero factor, since $ab = ac$ implies $a(b-c) = 0$ and hence $b = c$.

The additive group $\mathbb{Z}/n\mathbb{Z}$ becomes a ring with $\bar{a}\bar{b} = \overline{ab}$. This is well defined because replacing a, b by $a+nr, b+ns$ changes their product by $n(as+br+nrs)$. The ring identities follow from those in $\mathbb{Z}$. In $\mathbb{Z}/6\mathbb{Z}$, the nonzero classes $\bar{2}, \bar{3}$ have product zero, so cancellation fails: $\bar{2}\bar{1} = \bar{2}\bar{4}$, although $\bar{1} \neq \bar{4}$.

Proposition 1.27. *For $n \geq 2$, $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ is a unit if and only if* $\gcd(a,n) = 1$.

Proof. If $ab \equiv 1 \pmod{n}$, then $ab + nc = 1$ for some integer c, so every common divisor of a and n divides 1. Conversely the Euclidean algorithm gives integers b, c with $ab + nc = 1$ whenever the gcd is 1; reducing modulo n gives an inverse. For example $5 \cdot 5 - 2 \cdot 12 = 1$, so $\bar{5}^{-1} = \bar{5}$ in $\mathbb{Z}/12\mathbb{Z}$. □

Thus $\mathbb{Z}/p\mathbb{Z}$ is a field for a prime p; we also write this field as $\mathbb{F}_p$. If n is composite, a factorization $n = ab$ with $1 < a, b < n$ supplies zero divisors.

The *characteristic* of a nonzero ring A is the least positive integer n for which $n \cdot 1_A = 0$, if one exists, and is 0 otherwise. Here $n \cdot a$ means adding a to itself n times; negative integers act by additive inverses. Thus $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ have characteristic 0, while $\mathbb{F}_p$ has characteristic p. The characteristic of a field is either 0 or prime: a least positive value $n = ab$ with $1 < a, b < n$ would give $(a \cdot 1)(b \cdot 1) = 0$ with both factors nonzero in a domain. From now on, k denotes a field unless stated otherwise.

1.8 Polynomials and quotient rings

Let A be a commutative ring. The *polynomial ring* $A[x]$ consists of finite formal sums

$$f = a_0 + a_1 x + \cdots + a_d x^d, \qquad a_j \in A.$$

The symbol x is an *indeterminate*: equality means equality of each coefficient, rather than equality after giving x a particular value. We add coefficients in the same position and multiply by distributivity with $x^i x^j = x^{i+j}$. Thus the coefficient of x^r in fg is $\sum_{i+j=r} a_i b_j$, where missing coefficients are zero. These rules make $A[x]$ a commutative ring. For example, the coefficient of x^r in either $(fg)h$ or $f(gh)$ is the same finite sum $\sum_{i+j+\ell=r} a_i b_j c_\ell$, which proves associativity of multiplication; distributivity follows by distributing the coefficient sums. The constant polynomials identify A with a subring of $A[x]$.

For a nonzero polynomial, its *degree* $\deg f$ is the largest index d with $a_d \neq 0$; a_d is its *leading coefficient*, and $a_d x^d$ its leading term. The polynomial is *monic* if $a_d = 1$. We leave the degree of zero undefined and state zero-remainder cases separately. If A is a domain and $f, g \neq 0$, the product of their leading coefficients is nonzero, so

$$\deg(fg) = \deg f + \deg g.$$

In particular $A[x]$ is a domain. Over a field k, the units of $k[x]$ are precisely the nonzero constants: an inverse would force the sum of two nonnegative degrees to be zero.

The ring $A[x,y]$ consists of finite sums $\sum_{i,j} a_{ij} x^i y^j$, with $xy = yx$. It can also be constructed as $(A[x])[y]$. More generally, $A[x_1, \ldots, x_n]$ is obtained by adjoining one indeterminate at a time; repeated application of the preceding argument shows that it is a domain when A is a domain. A product $x_1^{d_1} \cdots x_n^{d_n}$ is a *monomial*, and its total degree is $d_1 + \cdots + d_n$. The *total degree* of a nonzero polynomial is the largest total degree among its monomials with nonzero coefficients. For example, $3x^2 y + y^2$ over $\mathbb{Q}$ has total degree 3.

Evaluation at $c \in k$ means substituting c for x, giving $f(c) = \sum_i a_i c^i$. An element c is a *root* of f if $f(c) = 0$. Different polynomials can define the same function over a finite field: $x^p - x$ is a nonzero polynomial but vanishes at every element of $\mathbb{F}_p$. Indeed, the multiplicative group $\mathbb{F}_p^\times$ has $p-1$ elements, so Lagrange's theorem gives $a^{p-1} = 1$ for $a \neq 0$; the identity $a^p = a$ holds also for $a = 0$. A formal polynomial and the function obtained by evaluation must therefore be distinguished.

For $f, g \in A[x]$, we say that g *divides* f if $f = qg$ for some $q \in A[x]$. The next result is the polynomial counterpart of integer division.

Theorem 1.28 (Division algorithm)**.** *For $f, g \in k[x]$ with $g \neq 0$, there exist unique $q, r \in k[x]$ such that $f = qg + r$ and either $r = 0$ or $\deg r < \deg g$.*

Proof. If $f = 0$, take $q = r = 0$. If $f \neq 0$ and $\deg f < \deg g$, take $q = 0, r = f$. Otherwise let the leading terms be ax^m and bx^d with $m \geq d$. Since $b \neq 0$ lies in a field, subtract $(a/b)x^{m-d} g$ from f. The leading term cancels, so the degree decreases. Repeat until the current remainder is zero or has degree smaller than g; strictly decreasing nonnegative degrees force termination. The subtracted multiples give q, and the final polynomial gives r. For uniqueness, subtract two proposed divisions to obtain $(q - q')g = r' - r$. If $q - q' \neq 0$,

the left side is nonzero and has degree at least $\deg g$, while the right side is either zero or has smaller degree. Either possibility is a contradiction. □

Dividing by $x-c$ leaves a constant remainder. Evaluating at c shows this constant is $f(c)$. Hence $f(c)=0$ if and only if $x-c$ divides f.

Corollary 1.29 (A bound on the number of roots)**.** *A nonzero polynomial of degree d over a field has at most d distinct roots. Consequently, over an infinite field, a polynomial that vanishes at every element is the zero polynomial.*

Proof. A nonzero constant has no roots. For the induction step, if f has a root a, write $f=(x-a)g$ by the remainder calculation. Then g has degree $d-1$. Every other root $b\neq a$ satisfies $0=(b-a)g(b)$, so $g(b)=0$ because $b-a$ is invertible in the field. By induction there are at most $d-1$ such roots. Including a gives at most d. If there is no root, the bound holds immediately. A finite degree cannot bound the size of an infinite set of roots, which proves the final assertion. □

A field k is *algebraically closed* if every nonconstant polynomial in $k[x]$ has a root in k. A *linear factor* is a polynomial of degree one, such as $x-c$. Repeated division by a linear factor shows that this is equivalent to every nonconstant polynomial being a product of linear factors and a nonzero constant. The fundamental theorem of algebra states that $\mathbb{C}$ is algebraically closed; we use this classical theorem without proof. The fields $\mathbb{R}$ and $\mathbb{Q}$ are not algebraically closed, since x^2+1 has no root in either.

An *ideal* $I\subseteq A$ is an additive subgroup such that $ai\in I$ for every $a\in A, i\in I$. Thus it contains zero, is closed under subtraction, and absorbs multiplication by arbitrary ring elements. The set of additive cosets is $A/I=\{a+I : a\in A\}$. Its addition is $(a+I)+(b+I)=(a+b)+I$, as for additive quotient groups. We write $a\equiv b \pmod I$ when $a-b\in I$, equivalently when their cosets are equal. The extra absorption condition permits multiplication on the quotient A/I. Specifically, if $a-a', b-b'\in I$, then

$$ab-a'b'=(a-a')b+a'(b-b')\in I.$$

Thus $(a+I)(b+I)=ab+I$ is independent of representatives. Merely being an additive subgroup would not prove this.

The *ideal generated by* $f_1,\dots,f_r$, denoted $(f_1,\dots,f_r)$, consists of sums $a_1f_1+\cdots+a_rf_r$ with $a_i\in A$. An ideal with such a finite generating list is *finitely generated*; an ideal generated by one element is *principal*. The zero ideal is $(0)=\{0\}$, the unit ideal is $(1)=A$, and an ideal is *proper* if it differs from A. For a subset $E\subseteq A$, the notation (E) means the ideal of all finite sums of multiples of elements of E. In $A/(f_1,\dots,f_r)$ each f_i becomes zero, and so does every ring-linear combination of the f_i. Quotient rings are a precise way of imposing equations.

Theorem 1.30 (First isomorphism theorem for rings)**.** *Let $\varphi: A\to B$ be a ring homomorphism. Its kernel is an ideal, its image is a subring of B, and*

$$A/\ker\varphi \longrightarrow \operatorname{im}\varphi, \qquad a+\ker\varphi \longmapsto \varphi(a)$$

is an isomorphism of rings.

Proof. The additive-group homomorphism property makes the kernel an additive subgroup. If $i \in \ker\varphi$ and $a \in A$, then $\varphi(ai) = \varphi(a)\varphi(i) = 0$, proving the absorption property. Images are closed under addition, negatives, and multiplication, and contain $\varphi(1) = 1$, so the image is a subring.

The displayed map is well defined because $a - a' \in \ker\varphi$ implies $\varphi(a) = \varphi(a')$. The original homomorphism identities show that it preserves sums, products, and the identity. It is onto the image by definition. If two classes have the same image, $\varphi(a - a') = 0$, so their difference belongs to the kernel and the classes are equal. Hence it is injective too. □

This is the group isomorphism argument with one extra check: the kernel must absorb multiplication so that a quotient *ring* exists. The parabola calculation below will be an explicit application.

Example 1.31 (Constructing a field by a relation)**.** In $\mathbb{R}[x]/(x^2+1)$ every element has a unique representative $a+bx$, by division. Sending $a+bx$ to $a+bi$ gives an isomorphism with $\mathbb{C}$: the equation $x^2 = -1$ makes multiplication agree. The uniqueness of the remainder proves injectivity, while every complex number has this form. Over $\mathbb{C}$ the same polynomial factors and the quotient is not a field; for example $(x-i)(x+i) = 0$ in that quotient and both factors have nonzero classes. The coefficient field matters.

Proposition 1.32 (Division and principal ideals)**.** *Every ideal of $\mathbb{Z}$ is generated by one integer. If k is a field, every ideal of $k[x]$ is generated by one polynomial.*

Proof. The zero ideal is generated by 0 in either ring. If $I \subseteq \mathbb{Z}$ is nonzero, it contains a positive integer; choose its least positive element d. For $a \in I$, integer division gives $a = qd + r$ with $0 \le r < d$. The remainder $r = a - qd$ lies in I, so minimality gives $r = 0$. Thus $I \subseteq (d)$, and the reverse inclusion follows because $d \in I$.

For a nonzero ideal $I \subseteq k[x]$, choose a nonzero $g \in I$ of smallest degree. For $f \in I$ divide $f = qg + r$. Since both f and qg lie in I, so does r. A nonzero r would have degree smaller than g, contradicting the choice. Thus $r = 0$, so $f \in (g)$. The reverse inclusion holds because $g \in I$ and ideals absorb multiplication. The two proofs use the same idea: a nonzero remainder would contradict a minimal choice. Polynomial division by an arbitrary nonzero polynomial is available here because its leading coefficient is invertible in the field k. □

1.9 A quotient ring associated to a parabola

A *polynomial map* $k^m \to k^n$ is a map whose n coordinate functions are polynomials in the m input coordinates. For example, consider

$$\gamma : k \longrightarrow k^2, \qquad t \longmapsto (t, t^2).$$

Its image is the set of solutions of $y = x^2$. To understand polynomials restricted to this set, substitute $x = t, y = t^2$:

$$\Phi : k[x,y] \longrightarrow k[t], \qquad f(x,y) \longmapsto f(t,t^2).$$

This is a ring homomorphism because substitution preserves sums, products, and 1. It is onto because $\Phi(x) = t$, and every polynomial $a_0 + a_1 t + \cdots + a_d t^d$ is the image of $a_0 + a_1 x + \cdots + a_d x^d$.

We compute its kernel without invoking algebraic geometry. Regard a polynomial in x, y as a polynomial in y with coefficients in $k[x]$. Division by the monic polynomial $y - x^2$ works over $k[x]$: cancel the highest power of y repeatedly, without dividing by a nonunit. This gives

$$f(x,y) = q(x,y)(y - x^2) + r(x).$$

Substitution gives $\Phi(f) = r(t)$. As a formal polynomial in t, this is zero if and only if every coefficient of r is zero. Consequently $\ker \Phi = (y - x^2)$, and

$$\boxed{k[x,y]/(y - x^2) \;\cong\; k[t].}$$

One can verify the isomorphism directly: send $\bar{x} \mapsto t, \bar{y} \mapsto t^2$ in one direction, and $t \mapsto \bar{x}$ in the other. The relation $\bar{y} = \bar{x}^2$ makes their composite the identity.

If k is infinite, this ring also records polynomial functions on the parabola: a polynomial in t vanishing for every $t \in k$ must be zero. If k is finite, the formal ring computation remains correct, but the ring of all polynomial functions on the finite point set has additional relations. Later the Nullstellensatz explains the algebra–geometry dictionary over an algebraically closed field.

The calculation contains the central idea of the course. Equations produce ideals; imposing equations produces quotient rings; and maps between geometric objects produce substitutions in the opposite direction. We will study these operations carefully before introducing more geometry.

Evaluating a polynomial and eliminating a variable

The parabola calculation illustrates a method that works with coefficients in any commutative ring A. Fix $g(x) \in A[x]$ and consider

$$\varepsilon : A[x,y] \longrightarrow A[x], \qquad F(x,y) \longmapsto F(x, g(x)).$$

For every $j \geq 1$,

$$y^j - g(x)^j = (y - g(x)) \sum_{i=0}^{j-1} y^{j-1-i} g(x)^i.$$

Writing $F(x,y) = \sum_j a_j(x) y^j$ and applying this identity term by term yields

$$F(x,y) - F(x,g(x)) \in (y - g(x)).$$

Consequently $\ker \varepsilon = (y - g(x))$: the containment from right to left follows by substitution, and the displayed difference proves the other containment whenever $F(x,g(x)) = 0$. Surjectivity follows because ε fixes $A[x]$. Hence

$$A[x,y]/(y-g(x)) \cong A[x].$$

No division by a coefficient was used. This is why elimination by a monic linear equation works even if A has zero divisors. It is also a useful model for later computations: define the map, verify its kernel, and only then identify the quotient.

1.10 Exercises

Exercise 1.1. Let D_4 be the symmetry group of a square, with r a quarter-turn and s a reflection, so $r^4 = s^2 = e$ and $sr = r^{-1}s$. Compute $(r^2s)(r^3s)$ and $(rs)^{-1}$.

Exercise 1.2 (Do the axioms really hold?)**.** Determine which of the following are groups under the specified operation: positive real numbers under multiplication; nonzero integers under multiplication; 2×2 real matrices under addition; invertible 2×2 real matrices under multiplication.

Exercise 1.3 (How many turns does a rotation need?)**.** Let g be an element of a group with finite order $n \geq 1$, and let $a \in \mathbb{Z}$. Prove that g^a has order $n/\gcd(n,a)$. Find the generators of the additive group $\mathbb{Z}/12\mathbb{Z}$.

Exercise 1.4. Find the kernel and image of the additive homomorphism $\mathbb{Z} \to \mathbb{Z}/12\mathbb{Z}$, $a \mapsto \overline{8a}$. Identify its quotient by the kernel.

Exercise 1.5 (The special case $p = 2$ without lists)**.** Show directly that a finite group of even order contains a nonidentity element g with $g^2 = e$.

Exercise 1.6 (Finding a generator by division)**.** Find a generator of the ideal $(x^3 - x, x^2 - 1) \subseteq k[x]$. Prove directly that the ideal $(x,y) \subseteq k[x,y]$ cannot be generated by one polynomial.

Exercise 1.7. Prove that every subgroup of the additive group $\mathbb{Z}$ is $d\mathbb{Z}$ for a unique integer $d \geq 0$.

Exercise 1.8. Explain why $\mathbb{Z} \subseteq \mathbb{Q}$ is an additive subgroup but not an ideal of $\mathbb{Q}$. What goes wrong with multiplication of additive cosets in $\mathbb{Q}/\mathbb{Z}$?

Exercise 1.9. In $k[\varepsilon]/(\varepsilon^2)$ write down all units, and explain why this ring is not a field. As usual in such a quotient, the same symbol ε denotes the residue class of the indeterminate; it satisfies $\varepsilon^2 = 0$.

Exercise 1.10. Compute $k[x,y]/(x,y)$ and $k[x,y]/(x)$. Explain why they are different even though both impose an equation involving x.

Lecture 2
Ideals, prime ideals, and the arithmetic of equations

We now study ideals systematically, using quotient rings to relate algebraic properties to equations. The first sections concern prime and maximal ideals, nilpotent elements, and simultaneous congruences. The later sections develop further ideal operations and the topology of the set of prime ideals. The lecture proves all numbered main-text results of [1, Chapter 1], including the three collections of identities labelled Exercises 1.12, 1.13, and 1.18 in that chapter.

2.1 Ideals and quotient rings

Throughout this course, rings are commutative, have an identity, and ring homomorphisms preserve that identity. The zero ring is allowed, but a field or an integral domain is required to be nonzero. An ideal $I \subseteq A$ is an additive subgroup such that $ai \in I$ whenever $a \in A$ and $i \in I$. As in Lecture 1, k denotes a field. An ideal is *proper* if $I \neq A$, or equivalently if $1 \notin I$: an ideal containing 1 contains every $a = a \cdot 1$.

The equations $f_1 = \cdots = f_r = 0$ imply every equation $\sum_i a_i f_i = 0$. Their ideal of algebraic consequences is

$$(f_1, \ldots, f_r) = \left\{ \sum_{i=1}^{r} a_i f_i : a_i \in A \right\}.$$

Thus ideal membership asks for an identity:

$$g \in (f_1, \ldots, f_r) \quad \Longleftrightarrow \quad g = \sum_{i=1}^{r} a_i f_i \quad \text{for some } a_i \in A.$$

For an infinite generating set, each such sum is still finite.

For two ideals, their *sum* is

$$I + J = \{ i + j : i \in I,\ j \in J \}.$$

It is an ideal: sums and differences stay in this set, and $a(i+j) = ai + aj$ does also. It contains both I and J, and every ideal containing both contains their sum. In particular, $I + (a) = \{i + ba : i \in I,\ b \in A\}$ is the ideal generated by I and the additional element a. This notation will be used in the maximal-ideal argument.

The quotient A/I consists of residue classes $a + I$. Its arithmetic is

$$(a+I) + (b+I) = (a+b) + I, \qquad (a+I)(b+I) = ab + I.$$

Independence of representatives follows from

$$(a+i)(b+j) - ab = aj + bi + ij \in I \qquad (i, j \in I).$$

The quotient map satisfies

$$q : A \longrightarrow A/I, \quad a \longmapsto a + I, \qquad \ker q = I, \quad \operatorname{im} q = A/I.$$

Example 2.1 (The equation of a parabola)**.** Define $\phi : k[x,y] \to k[t]$ by $x \mapsto t$ and $y \mapsto t^2$. It is surjective because t is in its image. Polynomial division in the variable y, by the monic polynomial $y - x^2$, gives

$$f(x,y) = h(x,y)(y - x^2) + r(x).$$

Consequently,

$$\phi(f) = 0 \iff r(t) = 0 \iff r = 0 \iff f \in (y - x^2).$$

Thus $\ker \phi = (y - x^2)$, and the first isomorphism theorem gives

$$k[x,y]/(y - x^2) \simeq k[t].$$

Thus the quotient is an ordinary polynomial ring in disguise. It describes polynomial expressions subject to the substitution rule $y = x^2$.

Proposition 2.2 (Ideals in a quotient; AM Proposition 1.1)**.** *Let I be an ideal of A. There is an inclusion-preserving bijection*

$$\{J \subseteq A : J \text{ an ideal},\ I \subseteq J\} \longleftrightarrow \{\bar{J} \subseteq A/I : \bar{J} \text{ an ideal}\},$$

with maps $J \mapsto J/I$ and $\bar{J} \mapsto q^{-1}(\bar{J})$. Here $J/I = q(J) = \{j + I : j \in J\}$. Inclusion-preserving means that $J_1 \subseteq J_2$ gives $J_1/I \subseteq J_2/I$, and conversely for the inverse correspondence.

Proof. For $j, j' \in J$ and $a \in A$,

$$(j+I) - (j'+I) = (j - j') + I \in J/I, \qquad (a+I)(j+I) = aj + I \in J/I.$$

Thus J/I is an ideal. For an ideal $\bar{J} \subseteq A/I$,

$$q(a), q(b) \in \bar{J} \implies q(a-b), q(ca) \in \bar{J} \qquad (c \in A),$$

so $q^{-1}(\bar{J})$ is an ideal containing $\ker q = I$. The inverse identities are

$$\begin{aligned} a \in q^{-1}(J/I) &\iff a - j \in I \text{ for some } j \in J \\ &\iff a \in J, \\ q\big(q^{-1}(\bar{J})\big) &= \bar{J} \qquad \text{since } q \text{ is onto.} \end{aligned}$$

Both image and inverse image preserve inclusions. □

Proposition 2.3 (Three ways to recognize a field; AM Proposition 1.2)**.** *For a nonzero ring A, the following are equivalent:*

(i) A is a field.
(ii) Its only ideals are (0) and A.
(iii) Every identity-preserving ring homomorphism $A \to B$ to a nonzero ring is injective.

Proof. (i)⇒(ii). For any nonzero ideal I, choose $0 \neq a \in I$. Then

$$1 = a^{-1}a \in I \quad \Longrightarrow \quad I = A.$$

(ii)⇒(iii). Since $f(1_A) = 1_B \neq 0_B$,

$$\ker f \neq A \quad \Longrightarrow \quad \ker f = 0 \quad \Longrightarrow \quad f \text{ is injective.}$$

(iii)⇒(i). If a is not a unit, apply (iii) to the quotient map $q : A \to A/(a)$:

$$(a) \neq A \implies A/(a) \neq 0 \implies (a) = \ker q = 0 \implies a = 0.$$

Thus every nonzero element is a unit. □

For example, an identity-preserving homomorphism from $\mathbb{Q}$ to any nonzero ring cannot identify two distinct rationals. A proposed homomorphism $\mathbb{Q} \to \mathbb{F}_p$ is impossible: it would send the nonzero rational integer p to zero, contradicting injectivity.

2.2 Prime and maximal ideals

Definition 2.4. A proper ideal $\mathfrak{p}$ is *prime* if $ab \in \mathfrak{p}$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. A proper ideal $\mathfrak{m}$ is *maximal* if no ideal lies strictly between $\mathfrak{m}$ and A. A nonzero ring is an *integral domain* if a product of two nonzero elements cannot be zero.

Theorem 2.5 (Quotient tests)**.** *For a proper ideal I,*

$$I \text{ prime} \iff A/I \text{ an integral domain}, \qquad I \text{ maximal} \iff A/I \text{ a field}.$$

Every maximal ideal is prime. The correspondence of Proposition 2.2 preserves prime and maximal ideals.

Proof. Write $\bar{a} = a + I$. The prime test follows from

$$\bar{a}\bar{b} = 0 \iff ab \in I, \qquad \bar{a} = 0 \text{ or } \bar{b} = 0 \iff a \in I \text{ or } b \in I.$$

If I is maximal, every $a \notin I$ satisfies

$$I + (a) = A \implies 1 = i + ba \quad (i \in I, b \in A) \implies \bar{b}\bar{a} = 1.$$

Hence A/I is a field. Conversely, suppose A/I is a field and $I \subsetneq J$. Choose $a \in J \setminus I$ and $b \in A$ with $\bar{a}\bar{b} = 1$. Then

$$ab \in J, \quad ab - 1 \in I \subseteq J \quad \implies \quad 1 = ab - (ab - 1) \in J,$$

so $J = A$.

A field is a domain: multiply $ab = 0$ by a^{-1} if $a \neq 0$. Finally, for $I \subseteq J$, the homomorphism $A/I \to A/J$, $a + I \mapsto a + J$, is well defined and surjective, with kernel J/I. The first isomorphism theorem therefore gives $(A/I)/(J/I) \simeq A/J$. A map obtained by passing a well-defined rule to residue classes is called an *induced map*. Applying the two quotient tests proves the last assertion. □

Example 2.6 (Prime and maximal ideals in familiar rings)**.**

ideal	quotient	conclusion
$(0) \subset \mathbb{Z}$	$\mathbb{Z}$	prime, not maximal
$(5) \subset \mathbb{Z}$	$\mathbb{F}_5$	maximal
$(x) \subset k[x,y]$	$k[y]$	prime, not maximal
$(x,y) \subset k[x,y]$	k	maximal
$(xy) \subset k[x,y]$	$k[x,y]/(xy)$	not prime
$(x^2+1) \subset \mathbb{R}[x]$	$\mathbb{C}$	maximal

For (xy), the nonzero residue classes of x and y have zero product. For the last entry, send x to i and divide by $x^2 + 1$ to see that the kernel is exactly $(x^2 + 1)$. Over $\mathbb{C}$, the same polynomial factors as $(x - i)(x + i)$, so the ideal it generates in $\mathbb{C}[x]$ is not prime. The field of coefficients matters.

2.2.1 Existence and the meaning of a local ring

A *partial order* on a set P is a relation $\leq$ that is reflexive and transitive and is *antisymmetric*: $x \leq y$ and $y \leq x$ imply $x = y$. A *chain* is a subset in which every two members are comparable. An *upper bound* for a subset T is an element $u \in P$ with $t \leq u$ for all $t \in T$. An element $m \in P$ is *maximal* if $m \leq x$ forces $x = m$. It need not be greater than every other element, since some elements may be incomparable.

We use *Zorn's lemma* as a foundational set-theoretic principle: a nonempty partially ordered set has a maximal element if every chain has an upper bound in the set. In the

application below, the elements of P are proper ideals and the order is inclusion. Every ring-theoretic step of that application will be proved.

Theorem 2.7 (Existence of maximal ideals; AM Theorem 1.3)**.** *Every nonzero ring has a maximal ideal.*

Proof. Order the proper ideals by inclusion. This collection contains (0). The union of a nonempty chain is an ideal: any two of its elements belong together to one member of the chain, where their difference can be taken; multiplication by an arbitrary ring element also stays in that member. The union cannot contain 1, since this would put 1 in a member of the chain. It is therefore an upper bound that is still proper. The empty chain has (0) as an upper bound. Zorn's lemma supplies a maximal member, which is a maximal ideal. □

Corollary 2.8 (AM Corollary 1.4)**.** *Every proper ideal I is contained in a maximal ideal.*

Proof. The ring A/I is nonzero, so Theorem 2.7 gives a maximal ideal in it. Its inverse image is a maximal ideal of A containing I, by the quotient correspondence and quotient tests. □

Corollary 2.9 (AM Corollary 1.5)**.** *Every nonunit belongs to a maximal ideal. Equivalently, the set of nonunits is the union of the maximal ideals.*

Proof. If a is not a unit, then (a) is proper, so use Corollary 2.8. Conversely, a proper ideal cannot contain a unit: multiplying by its inverse would put 1 in the ideal. □

Definition 2.10. A *local ring* is a ring with exactly one maximal ideal $\mathfrak{m}$. Its *residue field* is $A/\mathfrak{m}$. A ring with finitely many maximal ideals is called *semilocal*.

Proposition 2.11 (Two local-ring criteria; AM Proposition 1.6)**.** *(i) If $\mathfrak{m}$ is a proper ideal and every element of $A \setminus \mathfrak{m}$ is a unit, then A is local with maximal ideal $\mathfrak{m}$.*

(ii) If $\mathfrak{m}$ is maximal and every $1+t$, with $t \in \mathfrak{m}$, is a unit, then A is local.

Proof. For (i), any proper ideal consists of nonunits, hence is contained in $\mathfrak{m}$. Therefore $\mathfrak{m}$ is maximal and every maximal ideal equals it.

For (ii), take $x \notin \mathfrak{m}$. Maximality gives $xy + t = 1$ for some $y \in A$, $t \in \mathfrak{m}$. The product $xy = 1 - t$ is a unit by assumption. If u is its inverse, then $x(yu) = 1$, so x itself is a unit. Part (i) now applies. □

Example 2.12 (Local rings with explicit units)**.** In $A = k[\varepsilon]/(\varepsilon^2)$ every element has a unique form $a + b\varepsilon$. If $a \neq 0$, its inverse is

$$(a + b\varepsilon)^{-1} = a^{-1} - a^{-2}b\varepsilon.$$

If $a = 0$, it is nilpotent and cannot be a unit. Hence A is local with maximal ideal (ε) and residue field k. Similarly $\mathbb{Z}/8\mathbb{Z}$ is local, with maximal ideal $(\bar{2})$: the odd classes are exactly the units. The ring $\mathbb{Z}/6\mathbb{Z}$ has two maximal ideals, $(\bar{2})$ and $(\bar{3})$, so it is semilocal but not local.

2.3 Nilpotent elements and radicals

Definition 2.13. An element a is *nilpotent* if $a^n = 0$ for some $n \geq 1$. The set of nilpotents, called the *nilradical*, is denoted $\mathfrak{N}_A$. A ring with no nonzero nilpotents is *reduced*. The radical of an ideal I is

$$\sqrt{I} = \{a \in A : a^n \in I \text{ for some } n \geq 1\}.$$

An ideal is *radical* if $I = \sqrt{I}$.

Proposition 2.14 (Nilradical and the reduced quotient; AM Proposition 1.7)**.** *The nilpotent elements form an ideal $\mathfrak{N}_A$, and $A/\mathfrak{N}_A$ is reduced.*

Proof. If $a^m = 0$, then $(ca)^m = 0$ and $(-a)^m = 0$. If also $b^n = 0$, expand $(a+b)^{m+n-1}$ by distributivity. The binomial formula groups the terms according to how many factors contribute a and how many contribute b; the integer coefficients mean repeated addition in the ring. Every term contains $a^r b^s$ with $r+s = m+n-1$, so either $r \geq m$ or $s \geq n$. Every term is zero. Thus nilpotents are closed under addition and multiplication by arbitrary ring elements, and form an ideal.

Suppose the class of a in $A/\mathfrak{N}_A$ is nilpotent. Then $a^r \in \mathfrak{N}_A$ for some r, so $(a^r)^s = 0$ for some s. Hence $a \in \mathfrak{N}_A$, and its class was zero. This proves reducedness. □

Applying this to A/I shows that $\sqrt{I}$ is an ideal: it is the inverse image of $\mathfrak{N}_{A/I}$. It also shows

$$A/I \text{ is reduced} \quad \Longleftrightarrow \quad I = \sqrt{I}.$$

The rings $k[x]/(x)$ and $k[x]/(x^2)$ both have the single ordinary solution $x = 0$ over any field, but only the first is reduced. The second remembers more than the set of solutions, through its nonzero nilpotent $\bar{x}$.

The *prime spectrum* $\operatorname{Spec} A$ is, for the moment, simply the set of prime ideals of A. Thus an expression indexed by $\mathfrak{p} \in \operatorname{Spec} A$ ranges over all prime ideals. When intersecting ideals, the intersection of an empty family is taken to be the whole ring A.

Proposition 2.15 (Nilpotents are detected by all primes; AM Proposition 1.8)**.** *The nilradical is the intersection of all prime ideals:*

$$\mathfrak{N}_A = \bigcap_{\mathfrak{p} \in \operatorname{Spec} A} \mathfrak{p}.$$

In particular, an element lies in every prime ideal precisely when it is nilpotent.

Proof. If $a^n = 0 \in \mathfrak{p}$, repeated use of primality implies $a \in \mathfrak{p}$. Thus every nilpotent lies in every prime.

Conversely, suppose a is not nilpotent. The powers $S = \{1, a, a^2, \ldots\}$ avoid (0). Consider ideals disjoint from S. As in Theorem 2.7, the union of a chain is an ideal disjoint from S, since otherwise one member would already meet S. Zorn's lemma gives an ideal P maximal among those disjoint from S.

It is proper because $1 \in S$. If $u, v \notin P$, maximality forces $P + (u)$ and $P + (v)$ to meet S. Write

$$a^r = p + bu, \qquad a^s = q + cv \qquad (p, q \in P).$$

Multiplying yields $a^{r+s} \in P + (uv)$. If $uv \in P$, then P meets S, a contradiction. Thus $uv \notin P$; this is the contrapositive of primality. We have found a prime P which does not contain a. Consequently an element lying in every prime must be nilpotent. □

Definition 2.16. The *Jacobson radical* is

$$\mathrm{Jac}(A) = \bigcap_{\mathfrak{m} \text{ maximal}} \mathfrak{m}.$$

Since every maximal ideal is prime, $\mathfrak{N}_A \subseteq \mathrm{Jac}(A)$. The reverse inclusion need not hold.

Proposition 2.17 (A unit test for the Jacobson radical; AM Proposition 1.9)**.** *An element a belongs to* $\mathrm{Jac}(A)$ *if and only if* $1 - ab$ *is a unit for every* $b \in A$.

Proof. If $a \in \mathrm{Jac}(A)$ and $1 - ab$ is a nonunit, Corollary 2.9 supplies a maximal ideal $\mathfrak{m}$ with

$$1 - ab \in \mathfrak{m}, \quad ab \in \mathfrak{m} \quad \Longrightarrow \quad 1 = (1 - ab) + ab \in \mathfrak{m},$$

a contradiction. Conversely, if $a \notin \mathfrak{m}$ for some maximal ideal, invert its class in $A/\mathfrak{m}$:

$$\bar{a}\bar{b} = 1 \quad \Longrightarrow \quad 1 - ab \in \mathfrak{m} \quad \Longrightarrow \quad 1 - ab \text{ is a nonunit.}$$

□

Example 2.18 (The two radicals can differ)**.** Let $A = k[[t]]$, the *formal power-series ring*. An element is a sequence of coefficients written $\sum_{n \geq 0} a_n t^n$, with no requirement that only finitely many are nonzero. Equality and addition are coefficientwise; multiplication is defined by the coefficient formula

$$\left(\sum_{i \geq 0} a_i t^i \right) \left(\sum_{j \geq 0} b_j t^j \right) = \sum_{n \geq 0} \left(\sum_{i=0}^{n} a_i b_{n-i} \right) t^n.$$

Each output coefficient uses a finite sum, so this definition requires no analytic convergence. A series with $a_0 \neq 0$ has an inverse found successively from the coefficient equations for $fg = 1$: first $b_0 = a_0^{-1}$, then

$$b_n = -a_0^{-1} \sum_{i=1}^{n} a_i b_{n-i}.$$

A series with zero constant coefficient cannot be a unit. Thus A is local with maximal ideal (t) and $\mathrm{Jac}(A) = (t)$. It is a domain: the first nonzero coefficient of a product is the product of the first nonzero coefficients of its factors. Hence $\mathfrak{N}_A = 0$. In particular, t lies in the Jacobson radical but is not nilpotent.

2.4 Ideal operations and the Chinese remainder theorem

For ideals $I, J \subseteq A$, define

$$I+J = \{i+j : i \in I, j \in J\}, \qquad IJ = \left\{ \sum_{\nu=1}^{r} i_\nu j_\nu : i_\nu \in I, j_\nu \in J \right\}.$$

The sum is the smallest ideal containing both ideals. Their intersection is the largest ideal contained in both. The product is generated by all products of one element from each. In particular,

$$IJ \subseteq I \cap J, \qquad I(J+K) = IJ + IK.$$

For arbitrary families, sums still consist of finite sums of elements from the members. We define $I^0 = A$ and $I^n = I \cdots I$ with n factors for $n > 0$.

Example 2.19 (Do not confuse sum, intersection, and product)**.** In $\mathbb{Z}$,

$$(4)+(6) = (2), \qquad (4) \cap (6) = (12), \qquad (4)(6) = (24).$$

The first uses $2 = 6-4$, the second uses the least positive integer divisible by both 4 and 6, and the third ordinary multiplication. In $k[x,y]$,

$$(x,y)^2 = (x^2, xy, y^2).$$

Expansion proves that every product belongs to the displayed ideal, and every displayed generator is such a product. Ideals need not satisfy the distributive laws of subsets: with $I = (x)$, $J = (y)$, $K = (x+y)$, the element x lies in $I \cap (J+K)$ but not in $(I \cap J) + (I \cap K) = (xy, x(x+y))$, whose elements have no term of total degree one. For the first intersection, a polynomial divisible by both x and y has every monomial divisible by xy. For the second, if xh is divisible by $x+y$, substitution $y = -x$ gives $xh(x,-x) = 0$ in the domain $k[x]$. Thus $h(x,-x) = 0$, and division by the monic polynomial $y+x$ shows that h is divisible by $x+y$. These arguments justify the two intersection formulas.

Two ideals are *comaximal* if $I + J = A$, equivalently if $i + j = 1$ for some $i \in I$, $j \in J$. This is the ideal-theoretic version of relatively prime integers.

For rings $B_1, \ldots, B_r$, their *direct product* $\prod_i B_i$ is the set of tuples $(b_1, \ldots, b_r)$, with coordinatewise addition and multiplication. Its zero is $(0, \ldots, 0)$ and its identity is $(1, \ldots, 1)$. For example, in $k \times k$ one has $(1,0)(0,1) = (0,0)$. A product of rings is therefore different from the product of ideals inside a fixed ring; the meaning of $\prod$ is determined by its factors.

Proposition 2.20 (Chinese remainder theorem; AM Proposition 1.10)**.** *Let $I_1, \ldots, I_r$ be ideals, $r \geq 1$, and let*

$$\phi : A \longrightarrow \prod_{i=1}^{r} A/I_i, \qquad a \longmapsto (a+I_1, \ldots, a+I_r).$$

Then:

(i) If the ideals are pairwise comaximal, then $\prod_i I_i = \bigcap_i I_i$.

(ii) The map ϕ is surjective if and only if the ideals are pairwise comaximal.

(iii) The map ϕ is injective if and only if $\bigcap_i I_i = (0)$.

In particular, pairwise comaximal ideals give a ring isomorphism

$$A/(I_1 \cdots I_r) \simeq \prod_{i=1}^{r} A/I_i.$$

Proof. For $I + J = A$, choose $i \in I$, $j \in J$ with $i + j = 1$. Then

$$a \in I \cap J \quad \Longrightarrow \quad a = ai + aj \in JI + IJ = IJ.$$

Together with $IJ \subseteq I \cap J$, this gives equality. For induction, choose $u_i \in I_i$ with $u_i \equiv 1 \pmod{I_r}$, for $i < r$. Then

$$u = \prod_{i<r} u_i \in \prod_{i<r} I_i, \qquad 1 - u \in I_r,$$

so $(\prod_{i<r} I_i) + I_r = A$. Apply the two-ideal case and induction to prove (i).

If ϕ is surjective, for each pair $i \neq j$ there is an element a whose ith coordinate is 1 and jth coordinate is 0. Then $1 - a \in I_i$ and $a \in I_j$, so $1 \in I_i + I_j$.

Conversely, choose $u_{ij} \in I_j$ with $u_{ij} \equiv 1 \pmod{I_i}$ for each $j \neq i$. Put $e_i = \prod_{j \neq i} u_{ij}$, with empty product 1. Then

$$e_i \equiv \begin{cases} 1 & \pmod{I_i}, \\ 0 & \pmod{I_j}, \quad j \neq i, \end{cases} \qquad \phi\left(\sum_i a_i e_i\right) = (a_j + I_j)_j.$$

This proves surjectivity in (ii). Finally,

$$a \in \ker \phi \iff a \in I_i \text{ for every } i \iff a \in \bigcap_i I_i,$$

proving (iii). The first isomorphism theorem now gives the asserted isomorphism. □

2.5 Avoidance, quotients, and radical identities

Proposition 2.21 (Finite prime avoidance; AM Proposition 1.11)**.** *Let $r \geq 1$.*

(i) If an ideal I is contained in a finite union $\mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_r$ of prime ideals, then $I \subseteq \mathfrak{p}_i$ for some i.

(ii) If a prime ideal $\mathfrak{p}$ contains a finite intersection $I_1 \cap \cdots \cap I_r$, then it contains some I_i. If $\mathfrak{p} = \bigcap_i I_i$, then $\mathfrak{p} = I_i$ for some i.

Proof. For (i), prove the contrapositive by induction on r. For $r = 1$ there is nothing to prove. Suppose I is contained in none of the $\mathfrak{p}_i$. By induction, for each i choose $x_i \in I$ outside all $\mathfrak{p}_j$ with $j \neq i$. If one x_i also avoids $\mathfrak{p}_i$, it avoids their entire union and we are

done. Otherwise $x_i \in \mathfrak{p}_i$ for every i. Since $r \geq 2$, the element

$$y = \sum_{i=1}^{r} \prod_{j \neq i} x_j$$

belongs to I. Modulo $\mathfrak{p}_i$, every summand except the ith is zero because it contains x_i. The ith summand is not in $\mathfrak{p}_i$, since all its factors avoid that prime. Therefore $y \notin \mathfrak{p}_i$ for any i, proving the contrapositive.

For (ii), if no I_i is contained in $\mathfrak{p}$, choose $x_i \in I_i \setminus \mathfrak{p}$. The product $x_1 \cdots x_r$ lies in every I_i but not in $\mathfrak{p}$, a contradiction. In the equality case, the ideal I_i supplied by the first assertion satisfies both $I_i \subseteq \mathfrak{p}$ and $\mathfrak{p} \subseteq I_i$. □

Example 2.22 (Why the hypotheses of prime avoidance matter)**.** To see why the family must be finite, let $A = k[x_1, x_2, \ldots]$. This is the ring of polynomials in countably many variables, where each individual polynomial uses only finitely many variables. Put

$$I = (x_1, x_2, \ldots), \qquad \mathfrak{p}_n = (x_1, \ldots, x_n).$$

Each $\mathfrak{p}_n$ is prime because $A/\mathfrak{p}_n \simeq k[x_{n+1}, x_{n+2}, \ldots]$ is a domain: any proposed zero-divisor equation involves only finitely many variables, and a polynomial ring in those variables is a domain. Every element of I lies in some $\mathfrak{p}_n$, so $I = \bigcup_{n \geq 1} \mathfrak{p}_n$. Nevertheless I is contained in none of them, since $x_{n+1} \notin \mathfrak{p}_n$.

Primality also matters, even for a finite family. In $B = \mathbb{F}_2[x,y]/(x,y)^2$, write $\bar{x}, \bar{y}$ for the residue classes. The ideal $\mathfrak{m} = (\bar{x}, \bar{y})$ consists of the four elements $0, \bar{x}, \bar{y}, \bar{x} + \bar{y}$. The three subsets $\{0, \bar{x}\}$, $\{0, \bar{y}\}$, and $\{0, \bar{x} + \bar{y}\}$ are ideals. Each is closed under addition, and multiplying its nonzero member by $c + u$, with $c \in \mathbb{F}_2$ and $u \in \mathfrak{m}$, gives either that member or zero because $\mathfrak{m}^2 = 0$. They cover $\mathfrak{m}$, but none contains it. These three ideals are not prime: for example, $B/(\bar{x}) \simeq \mathbb{F}_2[y]/(y^2)$ is not a domain, and the same calculation applies to the other two.

Definition 2.23. The *ideal quotient* and *annihilator* are

$$(I : J) = \{a \in A : aJ \subseteq I\}, \qquad \operatorname{Ann}(J) = (0 : J).$$

We write $(I : x)$ for $(I : (x))$ and $\operatorname{Ann}(x) = (0 : x)$ for the annihilator of an element. These are ideals: the defining containment is preserved under subtraction and multiplication by any element of A.

Example 2.24. In $\mathbb{Z}$, $(12 : 8) = (3)$, since $12 \mid 8a$ exactly when $3 \mid a$. More generally $(m : n) = (m/\gcd(m,n))$ for positive m, n. In $k[x,y]$, $(xy : x) = (y)$: the equation $ax = xyb$ implies $a = yb$ by cancellation in the domain. Ideal quotients are a way to divide an ideal by a specified equation.

Proposition 2.25 (Ideal-quotient identities; AM main-text Exercise 1.12)**.** *For ideals I, J, K and arbitrary families of ideals,*

$$\begin{aligned}
&\textit{(i)}\ I \subseteq (I:J), \qquad \textit{(ii)}\ (I:J)J \subseteq I,\\
&\textit{(iii)}\ ((I:J):K) = (I:JK) = ((I:K):J),\\
&\textit{(iv)}\ \left(\bigcap_\lambda I_\lambda : J\right) = \bigcap_\lambda (I_\lambda : J),\\
&\textit{(v)}\ \left(I : \sum_\lambda J_\lambda\right) = \bigcap_\lambda (I : J_\lambda).
\end{aligned}$$

Proof. For (i)–(ii),

$$a \in I \implies aJ \subseteq I, \qquad a \in (I:J),\ j \in J \implies aj \in I.$$

Finite sums of the products aj still belong to I. For (iii),

$$\begin{aligned}
a \in ((I:J):K) &\iff akj \in I \quad (k \in K, j \in J)\\
&\iff aJK \subseteq I\\
&\iff a \in (I:JK).
\end{aligned}$$

Since $JK = KJ$, interchanging J, K gives the other equality. The remaining assertions follow from

$$\begin{aligned}
aJ \subseteq \bigcap_\lambda I_\lambda &\iff aJ \subseteq I_\lambda \quad \text{for every } \lambda,\\
a\sum_\lambda J_\lambda \subseteq I &\iff aJ_\lambda \subseteq I \quad \text{for every } \lambda.
\end{aligned}$$

In the last line, distribute a over each finite sum defining $\sum_\lambda J_\lambda$. □

Proposition 2.26 (Radical identities; AM main-text Exercise 1.13)**.** *For ideals I, J and a prime ideal $\mathfrak{p}$,*

$$\begin{aligned}
&\textit{(i)}\ I \subseteq \sqrt{I}, \qquad \textit{(ii)}\ \sqrt{\sqrt{I}} = \sqrt{I},\\
&\textit{(iii)}\ \sqrt{IJ} = \sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J},\\
&\textit{(iv)}\ \sqrt{I} = A \iff I = A,\\
&\textit{(v)}\ \sqrt{I+J} = \sqrt{\sqrt{I} + \sqrt{J}}, \qquad \textit{(vi)}\ \sqrt{\mathfrak{p}^n} = \mathfrak{p} \quad (n \geq 1).
\end{aligned}$$

Proof. For (i), use the first power. For (ii), if $a^r \in \sqrt{I}$ and $(a^r)^s \in I$, then $a^{rs} \in I$. The reverse inclusion follows from (i).

For (iii), $IJ \subseteq I \cap J$ gives one inclusion. If $a \in \sqrt{I} \cap \sqrt{J}$, choose $a^r \in I$ and $a^s \in J$. Then $a^{r+s} \in IJ$, giving the remaining inclusion around the chain. For (iv), if $1 \in \sqrt{I}$, then $1^n = 1 \in I$; the converse is immediate.

For (v), $I + J \subseteq \sqrt{I} + \sqrt{J}$ gives one inclusion after taking radicals. On the other hand, $\sqrt{I}$ and $\sqrt{J}$ both lie in the radical ideal $\sqrt{I+J}$; so their sum and then its radical lie in $\sqrt{I+J}$ by (ii).

For (vi), every element of $\mathfrak{p}$ has its nth power in $\mathfrak{p}^n$. Conversely, $a^r \in \mathfrak{p}^n \subseteq \mathfrak{p}$ implies $a \in \mathfrak{p}$ by primality. □

Proposition 2.27 (Radicals as intersections; AM Proposition 1.14)**.** *For every ideal I,*

$$\sqrt{I} = \bigcap_{\mathfrak{p} \supseteq I,\ \mathfrak{p}\ prime} \mathfrak{p}.$$

Proof. Let $q : A \to A/I$ be the quotient map. By Proposition 2.15 and the prime correspondence,

$$q^{-1}(\mathfrak{N}_{A/I}) = q^{-1}\left(\bigcap_{\mathfrak{p} \supseteq I} \mathfrak{p}/I\right) = \bigcap_{\mathfrak{p} \supseteq I} \mathfrak{p},$$
$$a \in q^{-1}(\mathfrak{N}_{A/I}) \iff q(a)^n = 0 \text{ for some } n \geq 1$$
$$\iff a \in \sqrt{I}.$$

The intersections range over prime ideals. If $I = A$, both sides are A by the empty-intersection convention. □

Example 2.28 (Computing radicals)**.** In $\mathbb{Z}$, $\sqrt{(12)} = (6)$: if $12 \mid a^n$, then 2 and 3 divide a; conversely $(6b)^2$ is divisible by 12. In $k[x,y]$,

$$\sqrt{(x^2, y^3)} = (x, y).$$

One inclusion follows because x^2, y^3 belong to the original ideal. The reverse follows because the maximal ideal (x,y) contains it. Also

$$\sqrt{(x^2 y)} = (xy) = (x) \cap (y).$$

Indeed, apply the product formula to $(x^2)(y)$ and note that $(x) \cap (y) = (xy)$, because a polynomial divisible by both x and y has every monomial divisible by xy. The two prime ideals describe the two coordinate axes; powers in the defining equation disappear on taking the radical.

Proposition 2.29 (Zero divisors and radical annihilators; AM Proposition 1.15)**.** *Let $D = \{a \in A : ab = 0 \text{ for some } b \neq 0\}$ be the set of zero divisors. Then*

$$D = \bigcup_{b \neq 0} \sqrt{\operatorname{Ann}(b)}.$$

Proof. If $ab = 0$ for $b \neq 0$, then $a \in \operatorname{Ann}(b) \subseteq \sqrt{\operatorname{Ann}(b)}$. Conversely, if $a \in \sqrt{\operatorname{Ann}(b)}$ with $b \neq 0$, then $a^n b = 0$ for some $n \geq 1$. Choose the smallest positive j for which $a^j b = 0$. The element $c = a^{j-1} b$ is nonzero by minimality, while $ac = 0$. Thus a is a zero divisor. This also proves directly the useful fact that if a power of an element is a zero divisor, then the element itself is a zero divisor. □

Proposition 2.30 (Comaximal radicals; AM Proposition 1.16)**.** *If $\sqrt{I}+\sqrt{J}=A$, then $I+J=A$.*

Proof. Write $u+v=1$ with $u^r \in I$, $v^s \in J$. Expand

$$1=(u+v)^{r+s-1}.$$

Every term contains either u^r or v^s, by the same exponent count used for nilpotents. Therefore every term belongs to $I+J$, so $1 \in I+J$. This also follows from parts (iv) and (v) of Proposition 2.26, but the expansion shows how to produce the required ideal membership. □

2.6 Extension and contraction of ideals

Let $f : A \to B$ be a ring homomorphism. For an ideal $I \subseteq A$, its image $f(I)$ need not be an ideal of B. We therefore define its *extension* by

$$I^e = Bf(I) = \left\{ \sum_{\nu=1}^{r} b_\nu f(a_\nu) : a_\nu \in I,\ b_\nu \in B \right\}.$$

For an ideal $J \subseteq B$, its *contraction* is the ideal $J^c = f^{-1}(J)$. Superscripts are read from left to right: I^{ec} means extend first, then contract. Both operations preserve inclusions. The defining formulas give the particularly useful equivalence

$$I^e \subseteq J \quad \Longleftrightarrow \quad I \subseteq J^c. \tag{2.1}$$

Indeed, the left condition is equivalent to $f(a) \in J$ for every $a \in I$, which is the condition on the right.

Example 2.31 (An extension can destroy primality)**.** For $\mathbb{Z} \hookrightarrow \mathbb{Q}$, the image $2\mathbb{Z}$ is not an ideal of $\mathbb{Q}$: it contains 2 but not $(1/2)2 = 1$. Its extension is $(2)^e = \mathbb{Q}$, although (2) was prime in $\mathbb{Z}$. For contraction, a prime always remains prime: if $ab \in f^{-1}(\mathfrak{q})$, then $f(a)f(b) \in \mathfrak{q}$, forcing one factor into $\mathfrak{q}$. Properness follows from $f(1) = 1$. Maximality need not remain: (0) is maximal in $\mathbb{Q}$, but its contraction (0) is not maximal in $\mathbb{Z}$.

Proposition 2.32 (Extension and contraction; AM Proposition 1.17)**.**

(i) $I \subseteq I^{ec}$ *and* $J^{ce} \subseteq J$.

(ii) $J^c = J^{cec}$ *and* $I^e = I^{ece}$.

(iii) An ideal I is contracted from B exactly when $I = I^{ec}$; an ideal J is extended from A exactly when $J = J^{ce}$. Extension and contraction give inverse bijections between the set of contracted ideals of A and the set of extended ideals of B.

Proof. The definitions give (i):

$$a \in I \implies f(a) \in I^e \implies a \in I^{ec}, \qquad J^{ce} = Bf(f^{-1}(J)) \subseteq J.$$

Apply these inclusions to J^c and I^e, and use preservation of inclusions:

$$J^c \subseteq J^{cec} \subseteq J^c, \qquad I^e \subseteq I^{ece} \subseteq I^e.$$

This proves (ii). For (iii),

$$\begin{aligned} I = J^c &\implies I^{ec} = J^{cec} = J^c = I, \\ I = I^{ec} &\implies I = (I^e)^c, \\ J = I^e &\implies J^{ce} = I^{ece} = I^e = J, \\ J = J^{ce} &\implies J = (J^c)^e. \end{aligned}$$

Thus extension and contraction restrict to inverse maps on the specified sets. □

Proposition 2.33 (Operations under change of ring; AM main-text Exercise 1.18)**.** *For ideals $I, I_1, I_2 \subseteq A$ and $J, J_1, J_2 \subseteq B$,*

$$\begin{aligned} (I_1+I_2)^e &= I_1^e + I_2^e, & (J_1+J_2)^c &\supseteq J_1^c + J_2^c, \\ (I_1 \cap I_2)^e &\subseteq I_1^e \cap I_2^e, & (J_1 \cap J_2)^c &= J_1^c \cap J_2^c, \\ (I_1 I_2)^e &= I_1^e I_2^e, & (J_1 J_2)^c &\supseteq J_1^c J_2^c, \\ (I_1 : I_2)^e &\subseteq (I_1^e : I_2^e), & (J_1 : J_2)^c &\subseteq (J_1^c : J_2^c), \\ (\sqrt{I})^e &\subseteq \sqrt{I^e}, & (\sqrt{J})^c &= \sqrt{J^c}. \end{aligned}$$

The extended ideals are closed under sums and products. The contracted ideals are closed under intersections, ideal quotients, and radicals.

Proof. We check the operations separately, because the directions of the inclusions matter.

Sums and intersections. The elements $f(a_1+a_2) = f(a_1) + f(a_2)$ generate the same ideal as the two sets $f(I_1)$ and $f(I_2)$ together, proving the sum equality. If $a_i \in J_i^c$, then $f(a_1+a_2) \in J_1 + J_2$, proving the contraction inclusion. Since $I_1 \cap I_2$ lies in both I_i, its extension lies in both extensions. Finally, $f(a)$ lies in both J_1, J_2 exactly when a lies in both inverse images.

Products. The generators $f(a_1a_2) = f(a_1)f(a_2)$ show that $(I_1I_2)^e$ and $I_1^e I_2^e$ coincide: expanding a product of finite generating sums proves the nontrivial containment. If $a_i \in J_i^c$, then $f(a_1a_2) \in J_1J_2$; sums of such products give the contraction inclusion.

Ideal quotients. If $aI_2 \subseteq I_1$, then $f(a)$ times any generator $f(b)$ of I_2^e lies in I_1^e. The same holds for every B-multiple and finite sum of the generators, so $f(a) \in (I_1^e : I_2^e)$. As the right side is an ideal, it contains the whole extension of $(I_1 : I_2)$. If $a \in (J_1 : J_2)^c$ and $b \in J_2^c$, then $f(a)f(b) \in J_1$, so $ab \in J_1^c$. This is the second quotient inclusion.

Radicals. If $a^n \in I$, then $f(a)^n \in I^e$. Thus each generator of $(\sqrt{I})^e$ lies in the radical ideal $\sqrt{I^e}$. For contraction, the exact equivalences

$$a \in (\sqrt{J})^c \iff f(a)^n \in J \text{ for some } n \iff a^n \in J^c \text{ for some } n$$

prove equality.

The sum and product formulas immediately prove closure for extended ideals. Intersections and radicals of contracted ideals are contracted by the equalities above. For the ideal quotient, if I_1, I_2 are contracted, we need a little more than the displayed inclusion. Directly from the definitions,

$$(I_1^e : I_2^e)^c = (I_1^{ec} : I_2) = (I_1 : I_2).$$

Indeed, $f(a)I_2^e \subseteq I_1^e$ is equivalent to $f(ab) \in I_1^e$ for every $b \in I_2$, and hence to $aI_2 \subseteq I_1^{ec}$. This exhibits $(I_1 : I_2)$ as a contraction and completes the closure assertion. □

Example 2.34 (Strict inclusions are real)**.** Under $k[t] \to k[x]$, $t \mapsto x^2$, the radical ideal (t) extends to (x^2), whose radical is (x). Thus $(\sqrt{(t)})^e \subsetneq \sqrt{(t)^e}$. Under the diagonal map $k \to k \times k$, the ideals $J_1 = k \times 0$ and $J_2 = 0 \times k$ both contract to zero, but $J_1 + J_2 = k \times k$ contracts to k. Thus contraction need not preserve sums.

2.7 Arithmetic of polynomial coefficients

An exercise at the end of this lecture relates polynomial arithmetic to properties of the coefficients. A polynomial $f = a_0 + \cdots + a_n x^n$ is called *primitive* here if its coefficients generate the unit ideal: $(a_0, \ldots, a_n) = A$. Over $\mathbb{Z}$, this says that their greatest common divisor is 1 up to sign.

Units and nilpotents in a polynomial ring

Here are complete coefficient arguments for two frequently used facts. For $f = a_0 + a_1 x + \cdots + a_d x^d \in A[x]$,

$$f \text{ is nilpotent} \iff a_0, \ldots, a_d \text{ are nilpotent},$$
$$f \text{ is a unit} \iff a_0 \text{ is a unit and } a_1, \ldots, a_d \text{ are nilpotent}.$$

For necessity in the first assertion, reduce modulo any prime $\mathfrak{p} \subseteq A$. The ring $(A/\mathfrak{p})[x]$ is a domain and has no nonzero nilpotents. Thus the image of every a_i is zero. Intersecting over all primes shows $a_i \in \sqrt{0}$.

For sufficiency, choose positive integers n_i with $a_i^{n_i} = 0$. Put $N = 1 + \sum_i (n_i - 1)$. Every product of N coefficients contains some a_i at least n_i times, so it vanishes. Expanding f^N therefore gives zero. Notice the use of a finite list of coefficients; the bound N is a uniform exponent for this list.

If $fg = 1$, the constant coefficients multiply to 1, so a_0 is a unit. Modulo any prime ideal, the degree formula for products in a domain shows that every unit polynomial is constant. Thus a_i belongs to every prime when $i > 0$, and is nilpotent. Conversely, if a_0 is a unit and the remaining coefficients are nilpotent, write $f = a_0(1 + h)$, where h is nilpotent by the first assertion. If $h^N = 0$, an inverse is

$$f^{-1} = a_0^{-1}(1 - h + h^2 - \cdots + (-h)^{N-1}).$$

For example, in $(\mathbb{Z}/8\mathbb{Z})[x]$,

$$(1+2x)^{-1} = 1 - 2x + 4x^2.$$

Multiplication gives $1 + 8x^3 = 1$. Positive polynomial degree therefore does not prevent invertibility over a ring with nilpotents.

Why a product of primitive polynomials is primitive

Suppose the coefficients of each of $f, g \in A[x]$ generate A. If the coefficients of fg generated a proper ideal, that ideal would lie in a maximal ideal $\mathfrak{m}$. Reducing modulo $\mathfrak{m}$, one would obtain $\bar{f}\bar{g} = 0$ in the domain $(A/\mathfrak{m})[x]$. Hence $\bar{f} = 0$ or $\bar{g} = 0$. All coefficients of one factor would then lie in $\mathfrak{m}$, contradicting primitivity. This proof works over arbitrary commutative rings and does not use a greatest common divisor. Over $\mathbb{Z}$ it is the primitive-product part of Gauss's lemma.

2.8 Boolean rings and idempotents

An element e is *idempotent* if $e^2 = e$. It behaves like a switch: multiplying twice by e does the same thing as multiplying once. The elements $(1,0)$ and $(0,1)$ in $k \times k$ are the basic examples. A *Boolean ring* is a ring in which every element is idempotent.

An idempotent gives a product decomposition

Let $e^2 = e$. The subsets eA and $(1-e)A$ are rings with respective identities e and $1-e$. Define

$$\Phi : A \longrightarrow eA \times (1-e)A, \qquad a \longmapsto (ea, (1-e)a).$$

This is a unital ring map because $e(1-e) = 0$. Its inverse is $(u,v) \mapsto u + v$: indeed $eu = u$, $ev = 0$, $(1-e)u = 0$, and $(1-e)v = v$. Thus

$$A \cong eA \times (1-e)A.$$

The decomposition has two nonzero factors exactly when $e \notin \{0,1\}$. It may equivalently be written $A \cong A/(1-e) \times A/(e)$ by the Chinese remainder theorem.

In a local ring every element outside the maximal ideal is a unit. At least one of $e, 1-e$ lies outside that ideal, since their sum is 1. If e is a unit, $e(1-e) = 0$ gives $e = 1$; if $1-e$ is a unit, it gives $e = 0$. A local ring therefore has no nontrivial product decomposition.

What the Boolean identity implies

In a Boolean ring, applying $a^2 = a$ to $a = 1+1$ gives $4 = 2$, hence $2 = 0$. Applying it to $a+b$ is then consistent with $(a+b)^2 = a+b$. Every quotient is Boolean. If $\mathfrak{p}$ is prime, the domain $A/\mathfrak{p}$ satisfies

$$\bar{a}(\bar{a}-1) = 0,$$

so each element is 0 or 1. Thus $A/\mathfrak{p} \cong \mathbf{F}_2$, and every prime ideal is maximal. Boolean rings are also reduced: $a^n = a$ for all $n \geq 1$, so a nilpotent element must be zero. These conclusions do not imply that a Boolean ring is finite; an arbitrary product of copies of $\mathbf{F}_2$ gives an infinite example.

2.9 The prime spectrum

A *topology* on a set can be specified by its closed subsets: arbitrary intersections and finite unions of closed subsets must be closed, and both the empty set and the whole set must be closed. Their complements are the *open* subsets. We now give $\operatorname{Spec} A$, the set of prime ideals of A, a topology using equations. A point of this space is an ideal, rather than initially a tuple of numbers.

For a subset $E \subseteq A$, put

$$V(E) = \{\mathfrak{p} \in \operatorname{Spec} A : E \subseteq \mathfrak{p}\}.$$

Read this as the set of points where all elements of E vanish. This is an algebraic definition; it does not require a prior geometric picture. The topology whose closed sets are the $V(E)$ is called the *Zariski topology*, and $\operatorname{Spec} A$ equipped with it is the *prime spectrum*. The closed-set axioms are verified in the exercises.

For $f \in A$, write

$$D(f) = \operatorname{Spec} A \setminus V(f) = \{\mathfrak{p} : f \notin \mathfrak{p}\}.$$

These are called *basic open sets*; the source uses the notation X_f. A family of opens is a *basis* if every open set is a union of members of the family. An *open cover* of a subset Y is a family of opens whose union contains Y; a *finite subcover* is a finite selection that still covers Y. A space is *quasi-compact* if every open cover of it has a finite subcover. A space is *Hausdorff* if any two distinct points have disjoint open neighborhoods. This additional separation property is not required for quasi-compactness.

For a subset $Y \subseteq X$ of a topological space, its *subspace topology* has opens $Y \cap U$ with U open in X; these are also called relatively open subsets of Y. Saying that Y is quasi-compact always refers to this topology. In particular, a cover of $D(f)$ may use sets $D(f) \cap U$ without requiring $U \subseteq D(f)$.

A point x is called a *closed point* if the one-element subset $\{x\}$ is closed. The *closure* of a subset is the intersection of all closed subsets containing it, hence is the smallest closed subset containing it.

Example 2.35 (An unusual but useful topology)**.** In $\operatorname{Spec}\mathbb{Z}$, the points are (0) and (p) for prime integers p. For a nonzero integer n, $V(n)$ consists of the finitely many prime ideals (p) with $p \mid n$. Thus $D(6)$ contains (0) and every (p) except (2) and (3). The point (0) is not closed: the smallest closed set containing it is the whole spectrum, because the only elements in (0) are zero. This is why familiar pictures of separated Euclidean points do not capture the spectrum topology.

A map $f : X \to Y$ of topological spaces is *continuous* if $f^{-1}(U)$ is open in X for every open $U \subseteq Y$. It is enough to check this for a basis of opens in Y, because inverse images preserve unions. A *homeomorphism* is a bijection for which both the map and its inverse are continuous. A space is *disconnected* if it is a union of two disjoint nonempty open subsets; otherwise it is *connected*. A subset which is both open and closed is sometimes called *clopen*.

The exercise on idempotents at the end of this lecture connects three descriptions of the same phenomenon: a geometric space splits into separate pieces, its ring of functions splits into a product, and an idempotent function is 1 on one piece and 0 on the other. For the interpolation ring $k[t]/(t(t-1))$, the two functions are precisely $1-t$ and t. In the next lecture, modules will let us study linear equations over these rings, including equations whose coefficients have zero divisors.

2.10 Exercises

Exercise 2.1 (Prime ideals of an integer quotient)**.** Find all prime and maximal ideals of $\mathbb{Z}/12\mathbb{Z}$.

Exercise 2.2 (Nilpotents can be added to units; AM Chapter 1, Exercise 1)**.** Let A be a ring and let $a \in A$ be nilpotent. Prove that $1+a$ is a unit of A. Deduce that a unit plus a nilpotent is again a unit.

Exercise 2.3 (Congruences become interpolation)**.** Solve $n \equiv 2 \pmod 3$, $n \equiv 4 \pmod 5$. Then identify $k[t]/(t(t-1))$.

Exercise 2.4 (Arithmetic of polynomial coefficients; AM Chapter 1, Exercise 2)**.** Let A be a ring, and let $f = a_0 + \cdots + a_n x^n$ and g belong to $A[x]$. Recall that a polynomial is primitive here when its coefficients generate the unit ideal. Prove:

(i) f is a unit exactly when a_0 is a unit and $a_1, \ldots, a_n$ are nilpotent.

(ii) f is nilpotent exactly when every coefficient is nilpotent.

(iii) f is a zero divisor exactly when there is a nonzero $a \in A$ with $af = 0$.

(iv) A product fg is primitive exactly when both f and g are primitive.

Exercise 2.5 (A polynomial ring has equal radicals; AM Chapter 1, Exercise 4)**.** For every ring A, prove $\operatorname{Jac}(A[x]) = \mathfrak{N}_{A[x]}$.

Exercise 2.6 (Boolean rings; AM Chapter 1, Exercise 11)**.** If A is Boolean, prove:

(i) $2a = 0$ for every $a \in A$.

(ii) Every prime ideal is maximal and its quotient is the two-element field.

(iii) Every finitely generated ideal is principal.

Exercise 2.7 (Idempotents in a local ring; AM Chapter 1, Exercise 12)**.** Prove that an idempotent in a local ring is either 0 or 1.

Exercise 2.8 (The Zariski closed sets; AM Chapter 1, Exercise 15)**.** Let A be a ring, let I, J be ideals of A, and let E, E_λ be subsets of A. Recall $V(E) = \{\mathfrak{p} \in \operatorname{Spec} A : E \subseteq \mathfrak{p}\}$. Prove the following identities, and deduce that the sets $V(E)$ are the closed sets of a topology:

(i) If $I = (E)$, then $V(E) = V(I) = V(\sqrt{I})$.

(ii) $V(0) = \operatorname{Spec} A$ and $V(1) = \varnothing$.

(iii) $V(\bigcup_\lambda E_\lambda) = \bigcap_\lambda V(E_\lambda)$.

(iv) $V(I \cap J) = V(IJ) = V(I) \cup V(J)$.

Exercise 2.9 (Basic opens and finiteness; AM Chapter 1, Exercise 17)**.** For a ring A, put $D(f) = \{\mathfrak{p} \in \operatorname{Spec} A : f \notin \mathfrak{p}\}$ for $f \in A$. Show that these sets form a basis for the Zariski topology on $\operatorname{Spec} A$, and prove the following assertions for $f, g \in A$:

(i) $D(f) \cap D(g) = D(fg)$.

(ii) $D(f) = \varnothing$ exactly when f is nilpotent.

(iii) $D(f) = \operatorname{Spec} A$ exactly when f is a unit.

(iv) $D(f) = D(g)$ exactly when $\sqrt{(f)} = \sqrt{(g)}$.

(v) $\operatorname{Spec} A$ is quasi-compact.

(vi) Every $D(f)$ is quasi-compact.

(vii) An open subset is quasi-compact exactly when it is a finite union of basic opens.

Exercise 2.10 (Idempotents and disconnected spaces; AM Chapter 1, Exercise 22)**.** For a finite product $A = \prod_{i=1}^r A_i$, show that $\operatorname{Spec} A$ is the disjoint union of open and closed subspaces naturally homeomorphic to $\operatorname{Spec} A_i$. Then prove that the following are equivalent for an arbitrary ring A:

(i) $\operatorname{Spec} A$ is disconnected, meaning a union of two disjoint nonempty open subsets.

(ii) $A \simeq A_1 \times A_2$ with both factors nonzero.

(iii) A has an idempotent different from 0 and 1.

Deduce that the spectrum of a local ring is connected.

Lecture 3
Modules: linear algebra over a ring

Modules extend the language of vector spaces to an arbitrary commutative ring. The examples of abelian groups and ideals will show why generators can satisfy relations and why a submodule of a free module need not be free. Throughout this lecture, A is a commutative ring with identity. Reading: [1, Chapter 2, sections on modules, submodules, and finitely generated modules].

3.1 Modules and module homomorphisms

Definition 3.1. An *A-module M* is an abelian group, written additively, together with scalar multiplication $A \times M \to M$, $(a,m) \mapsto am$, satisfying

$$(a+b)m = am+bm, \quad a(m+n) = am+an, \quad (ab)m = a(bm), \quad 1m = m.$$

An *A-linear map*, or *module homomorphism*, $f : M \to N$ satisfies $f(m+n) = f(m)+f(n)$ and $f(am) = af(m)$. It is an *isomorphism* if it has an A-linear inverse. An A-linear map $M \to M$ is an *endomorphism*.

No multiplication between two elements of M is required. This is why an ideal, a space of vectors, and an abelian group can all belong to the same theory. The equation $0m = 0$ follows by subtracting $0m$ from $(0+0)m = 0m+0m$. Similarly $a0 = 0$, and $(-a)m = -(am)$. For the zero ring, the rule $1m = m$ forces every module to be zero.

For $n \geq 1$, the module A^n consists of tuples $(a_1, \ldots, a_n)$ with coordinatewise addition and scalar multiplication. We set $A^0 = 0$, the module with one element.

A *submodule* $N \subseteq M$ is an additive subgroup stable under scalar multiplication: $an \in N$ whenever $a \in A$ and $n \in N$. Thus it is a module with the operations inherited from M. A bijective linear map is an isomorphism: if $f(m) = n$ and $f(m') = n'$, then $f(m+m') = n+n'$ and $f(am) = an$, so its inverse preserves addition and scalars.

Example 3.2 (Abelian groups, vector spaces, and ideals)**.** Every vector space over a field k is a k-module, and these notions are identical. Every abelian group G is a $\mathbb{Z}$-module: for $n > 0$, the element ng is the sum of n copies of g, and negative integers act by additive

inverses. This is the only possible $\mathbb{Z}$-module structure. Finally, every ideal $I \subseteq A$ is an A-module using the ring's multiplication. An ideal is therefore a submodule of the module A.

The group $\mathbb{Z}/6\mathbb{Z}$ is a module with a nonzero element $\bar{3}$ killed by the nonzero scalar 2. In a vector space this cannot happen: multiplying $2v = 0$ by 2^{-1} would imply $v = 0$ whenever $2 \neq 0$ in the field. Over $\mathbb{Z}$ there is no inverse scalar $1/2$.

Definition 3.3. For an element $m \in M$, its *annihilator* is

$$\mathrm{Ann}_A(m) = \{a \in A : am = 0\}.$$

The annihilator of the module is

$$\mathrm{Ann}_A(M) = \{a \in A : am = 0 \text{ for every } m \in M\}.$$

The module is *faithful* if $\mathrm{Ann}_A(M) = 0$.

Both annihilators are ideals. For example, if a, b kill m, then $(a-b)m = am - bm = 0$, and $(ra)m = r(am) = 0$ for every $r \in A$. The same argument applied to every $m \in M$ proves the assertion for $\mathrm{Ann}_A(M)$.

For instance, $\mathrm{Ann}_{\mathbb{Z}}(\mathbb{Z}/6\mathbb{Z}) = (6)$, whereas $\mathrm{Ann}_{\mathbb{Z}}(\bar{3}) = (2)$. Distinguish a scalar that kills one chosen element from a scalar that kills the entire module.

3.2 Submodules, quotients, and the first isomorphism theorem

Let $N \subseteq M$ be a submodule. Define scalar multiplication on the quotient group by

$$a(m+N) = am+N.$$

It is independent of the representative because

$$m - m' \in N \quad \Longrightarrow \quad am - am' = a(m-m') \in N.$$

Here closure under scalar multiplication is essential.

For submodules $N, P \subseteq M$, their *sum* and *intersection* are

$$N+P = \{n+p : n \in N,\ p \in P\}, \qquad N \cap P = \{m : m \in N \text{ and } m \in P\}.$$

Both are submodules, by closure under addition and scalars. More generally, the submodule generated by a subset $E \subseteq M$ consists of all finite sums $\sum_i a_i e_i$ with $a_i \in A$ and $e_i \in E$. For a family (N_i), the notation $\sum_i N_i$ means the submodule generated by their union; each individual element uses only finitely many summands.

For an ideal $I \subseteq A$, we write IM for the submodule of finite sums $\sum_i a_i m_i$ with $a_i \in I$ and $m_i \in M$. For instance, $2M = \{2m : m \in M\}$ for a $\mathbb{Z}$-module M.

For an A-linear map $f : M \to N$, define its *kernel*, *image*, and *cokernel* by

$$\ker f = \{m : f(m) = 0\}, \qquad \operatorname{im} f = \{f(m) : m \in M\}, \qquad \operatorname{coker} f = N/\operatorname{im} f.$$

The kernel and image are submodules. For example, if $m, m' \in \ker f$, then $f(am + bm') = af(m) + bf(m') = 0$.

Theorem 3.4 (First isomorphism theorem for modules)**.** *The rule*

$$M/\ker f \longrightarrow \operatorname{im} f, \qquad m + \ker f \longmapsto f(m)$$

is an A-module isomorphism.

Proof. Put $K = \ker f$ and $\bar{f}(m+K) = f(m)$. Then

$$m + K = m' + K \iff m - m' \in K \iff f(m) = f(m').$$

This proves both well-definedness and injectivity. Moreover,

$$\bar{f}(a(m+K) + b(n+K)) = af(m) + bf(n), \qquad \operatorname{im} \bar{f} = \operatorname{im} f.$$

Thus $\bar{f}$ is linear and surjective onto its stated target. □

The *quotient factorization property* is the equivalence

$$g(N) = 0 \quad \iff \quad g = \bar{g} \circ q \text{ for a unique linear } \bar{g} : M/N \to P,$$

where $q(m) = m + N$ and $\bar{g}(m+N) = g(m)$. Indeed,

$$m - m' \in N \implies g(m) - g(m') = g(m - m') = 0.$$

Conversely, $g = \bar{g}q$ vanishes on $N = \ker q$. Surjectivity of q gives uniqueness.

Example 3.5 (A matrix over the integers)**.** Consider $f : \mathbb{Z}^2 \to \mathbb{Z}$, $f(a,b) = 2a + 3b$. Since $f(-1,1) = 1$, this is surjective. If $2a + 3b = 0$, then b is even, say $b = 2t$, and then $a = -3t$. Thus

$$\ker f = \mathbb{Z}(-3,2), \qquad \mathbb{Z}^2/\mathbb{Z}(-3,2) \simeq \mathbb{Z}.$$

Here $\mathbb{Z}(-3,2) = \{t(-3,2) : t \in \mathbb{Z}\}$ is the submodule generated by $(-3,2)$. This resembles computing a matrix kernel over a field, but divisibility must be handled explicitly: an integer solution cannot be obtained by freely dividing by arbitrary nonzero integers.

3.3 Generators, bases, and presentations

A family $(m_i)_{i \in S}$ *generates* M if every element of M is a finite sum $\sum a_i m_i$. A module is *finitely generated* if some finite family generates it. In commutative algebra a *finite module* means a finitely generated module, not necessarily a set with finitely many elements. A module generated by one element is called *cyclic*. A generating family is a *basis* if those coefficients are unique. A module with a basis is *free*. A module with a basis of n elements

is called free of *rank n*; for a nonzero ring this number is independent of the chosen finite basis, by the corollary below. For example, A^n has the standard basis $e_1, \dots, e_n$, where e_i has 1 in its ith coordinate and 0 elsewhere.

Proposition 3.6 (Generators give a quotient of a free module; AM 2.3)**.** *Elements $m_1, \dots, m_n$ generate M if and only if the map*

$$\pi : A^n \longrightarrow M, \qquad (a_1, \dots, a_n) \longmapsto \sum_i a_i m_i$$

is surjective. Consequently, M is finitely generated if and only if it is isomorphic to a quotient of A^n for some finite n (including $n = 0$). Its kernel consists of all relations among these generators. They form a basis if and only if π is also injective.

Proof. By definition,

$$\operatorname{im}\pi = \sum_{i=1}^{n} A m_i, \qquad \ker\pi = \left\{ (a_i) \in A^n : \sum_i a_i m_i = 0 \right\}.$$

Moreover,

$$\sum_i a_i m_i = \sum_i b_i m_i \quad \Longleftrightarrow \quad (a_i - b_i)_i \in \ker\pi.$$

Thus generation is equivalent to surjectivity, and uniqueness of coefficients to $\ker\pi = 0$. The first isomorphism theorem gives $M \cong A^n/\ker\pi$ when π is onto. Conversely, for any submodule $K \subseteq A^n$,

$$(a_1, \dots, a_n) + K = \sum_i a_i (e_i + K),$$

so $e_1 + K, \dots, e_n + K$ generate A^n/K. □

Corollary 3.7 (A finite free module has a well-defined rank)**.** *If $A \neq 0$ and $A^r \simeq A^s$ as A-modules, then $r = s$.*

Proof. Choose a maximal ideal $\mathfrak{m}$ and put $k = A/\mathfrak{m}$. Linearity of f and f^{-1} gives

$$f(\mathfrak{m}A^r) = \mathfrak{m}A^s.$$

Hence f induces an isomorphism

$$k^r = A^r/\mathfrak{m}A^r \xrightarrow{\sim} A^s/\mathfrak{m}A^s = k^s.$$

Taking dimensions over k yields $r = s$. □

Definition 3.8 (Exact sequences)**.** A pair of A-linear maps $L \xrightarrow{u} M \xrightarrow{v} N$ is *exact at M* if $\operatorname{im} u = \ker v$. A sequence is *exact* if this condition holds at every displayed intermediate module. An exact sequence of the form

$$0 \longrightarrow L \xrightarrow{u} M \xrightarrow{v} N \longrightarrow 0$$

is called a *short exact sequence*. Here 0 is the zero module, and each arrow to or from it is the unique zero map. Thus exactness at L says that u is injective, and exactness at N says that v is surjective.

If the relation module $\ker\pi$ is generated by r vectors $v_1,\dots,v_r \in A^n$, arrange these vectors as the columns of an $n\times r$ matrix D. Then

$$M \simeq A^n/\operatorname{im} D.$$

This is a *finite presentation*: n symbols generate M, and the r columns specify all relations. We also write

$$A^r \xrightarrow{D} A^n \longrightarrow M \longrightarrow 0.$$

Exactness says that the image of D is precisely the kernel of the quotient map $A^n \to M$, and that this quotient map is surjective. The map D need not be injective: relations themselves may satisfy relations. We call M *finitely presented* if such a presentation exists with both r and n finite. Lecture 4 studies further consequences of exactness.

Example 3.9 (A cyclic module need not be free)**.** The module $\mathbb{Z}/6\mathbb{Z}$ is generated by $\bar{1}$ with relation $6\bar{1}=0$:

$$\mathbb{Z} \xrightarrow{\times 6} \mathbb{Z} \longrightarrow \mathbb{Z}/6\mathbb{Z} \longrightarrow 0.$$

It is not free over $\mathbb{Z}$. Indeed, every nonzero free $\mathbb{Z}$-module has no nonzero element killed by a nonzero integer, as one sees coordinate by coordinate; $\bar{1}$ is killed by 6. Also, the ideal $2\mathbb{Z}$ is free of rank one: the map $\mathbb{Z}\to 2\mathbb{Z}$, $n\mapsto 2n$, is an isomorphism of modules. This map is not an isomorphism of unital rings, and its inverse is a module map on $2\mathbb{Z}$, not multiplication by a scalar in $\mathbb{Z}$.

Example 3.10 (The ideal (x,y) has a relation)**.** Let $A=k[x,y]$, and consider $\pi: A^2\to(x,y)$, $(a,b)\mapsto ax+by$. The vector $(-y,x)$ belongs to the kernel. We claim it generates the whole kernel.

If $ax+by=0$, evaluation at $x=0$ and cancellation give

$$\begin{aligned} b(0,y)y=0 &\implies b(0,y)=0 \implies b=xc \quad (c\in A),\\ x(a+yc)=0 &\implies a=-yc. \end{aligned}$$

Here $k[y]$ and $A=k[x,y]$ are domains. Thus

$$(a,b)=c(-y,x), \qquad \ker\pi = A(-y,x),$$

proving

$$A \xrightarrow{\binom{-y}{x}} A^2 \longrightarrow (x,y) \longrightarrow 0.$$

In fact the first map is injective, because $cx=0$ implies $c=0$.

The ideal (x,y) is not principal. If $(x,y)=(f)$, then f divides both x and y. Since total degrees of nonzero polynomials add under multiplication, a nonconstant divisor of x has degree one and is a nonzero constant multiple of x. Such a polynomial does not divide y. Hence f would have to be a unit, forcing $(x,y)=A$, contrary to evaluation at $(0,0)$. In fact (x,y) is not free at all. Any two nonzero elements u,v of an ideal in a domain satisfy

the nontrivial module relation $vu - uv = 0$, with coefficients $v, -u$. Thus a basis of such an ideal could have at most one element. A one-element basis would make the ideal principal, which we have just ruled out. This gives a concrete submodule of the free module A that is not free.

Caution 3.11. Finitely generated and finitely presented are different conditions over a general ring. The first gives finitely many generators for M; the second also requires finitely many generators for the relation module of a suitable finite presentation. A ring is called *Noetherian* if every ideal is finitely generated. Lecture 9 studies this condition and proves that it makes every finitely generated module finitely presented. That implication requires a proof and is not part of the definition of a finitely generated module.

3.4 Polynomial modules and linear operators

Let V be a vector space over k, and let $T : V \to V$ be a linear operator, that is, a k-linear endomorphism. Write $T^0 = \mathrm{id}_V$ and $T^{i+1} = T \circ T^i$. Define a $k[t]$-module structure by

$$f(t) \cdot v = f(T)v, \qquad (a_0 + a_1 t + \cdots + a_r t^r) \cdot v = \sum_{i=0}^{r} a_i T^i v.$$

For $f = \sum_i a_i t^i$ and $g = \sum_j b_j t^j$, both $(fg)(T)$ and $f(T)g(T)$ equal $\sum_{i,j} a_i b_j T^{i+j}$, because $T^i T^j = T^{i+j}$. This proves associativity of scalar multiplication. Conversely, a $k[t]$-module is a k-vector space together with the linear operator $T(v) = tv$. Thus studying a linear operator is the same as studying a module over a polynomial ring in one variable.

A linear operator is *nilpotent* if $T^r = 0$ for some positive integer r. This definition concerns composition of maps, just as nilpotence in a ring concerns powers under multiplication.

A matrix becomes a module presentation

Let $V = k^2$ with basis v_1, v_2, and let $T(v_1) = 0$, $T(v_2) = v_1$. In the associated $k[t]$-module,

$$tv_2 = v_1, \qquad tv_1 = 0, \qquad t^2 v_2 = 0.$$

Thus v_2 alone generates the module over $k[t]$, even though two vectors are needed over k. The map $k[t] \to V$, $f \mapsto f(T)v_2$, is onto. Division by t^2 writes $f = qt^2 + a + bt$, and

$$f(T)v_2 = av_2 + bv_1.$$

Independence of v_1, v_2 shows that the kernel is exactly (t^2). Hence $V \cong k[t]/(t^2)$ as a $k[t]$-module.

Now compare the zero operator on k^2. Its associated module is $k[t]/(t) \oplus k[t]/(t)$, which is not cyclic: a polynomial acts as its constant term, so one vector generates only a one-

dimensional subspace. Both modules have k-dimension two, but their annihilators are (t^2) and (t), respectively. The module records the operator as well as the underlying vector space.

A map $S : V \to W$ between modules arising from operators T_V, T_W is $k[t]$-linear exactly when it is k-linear and

$$ST_V = T_W S.$$

Necessity follows by applying linearity to tv. For sufficiency, induction gives $ST_V^j = T_W^j S$, and then the same identity for every polynomial in the operators. This is the algebraic reason that invariant subspaces and commuting matrices enter the study of polynomial modules.

3.5 Computing maps of modules

The set of A-linear maps from M to N is written $\mathrm{Hom}_A(M,N)$. When $M = N$ we also write $\mathrm{End}_A(M)$. It is itself an A-module, with pointwise operations $(f+g)(m) = f(m)+g(m)$ and $(af)(m) = af(m)$. Commutativity of A ensures that $(af)(bm) = abf(m) = b(af)(m)$, so af remains linear.

A map from A is determined by the image of 1, giving

$$\mathrm{Hom}_A(A,N) \simeq N, \qquad f \longmapsto f(1).$$

More generally,

$$\mathrm{Hom}_A(A/I,N) \simeq \{n \in N : In = 0\}.$$

The inverse sends n to $f_n(a+I) = an$. Precisely,

$$In = 0, \quad a - a' \in I \quad \Longrightarrow \quad an = a'n,$$

so f_n is well defined, and

$$f_n(1+I) = n, \qquad f(a+I) = af(1+I).$$

These identities verify that the two maps are inverse.

Remark 3.12 (Torsion over a domain). When A is a domain, an element m is called *torsion* if $am = 0$ for some nonzero $a \in A$. Torsion elements form a submodule, denoted $T(M)$: if $am = 0$ and $bn = 0$, then $ab(m+n) = 0$, and $ab \neq 0$. A module is *torsion-free* if its only torsion element is zero. Free modules over a domain are torsion-free, but the converse requires additional hypotheses and is false in general. This vocabulary will help distinguish phenomena that ordinary vector-space dimension cannot see.

3.6 Isomorphism theorems and annihilators

The next two statements complete the main-text submodule results of [1, Chapter 2]. They are often used inside longer proofs, so it is useful to know the actual maps rather than only the names of the isomorphism theorems.

Proposition 3.13 (Two quotient identities; AM 2.1)**.**

1. *If $N \subseteq M \subseteq L$, then $(L/N)/(M/N) \cong L/M$.*
2. *If $M_1, M_2 \subseteq M$, then $(M_1 + M_2)/M_1 \cong M_2/(M_1 \cap M_2)$.*

Proof. The first map is

$$\alpha : L/N \longrightarrow L/M, \qquad x+N \longmapsto x+M.$$

The inclusion $N \subseteq M$ makes it well defined, and

$$\operatorname{im}\alpha = L/M, \qquad \ker\alpha = \{x+N : x \in M\} = M/N.$$

For the second map, set

$$\beta : M_2 \longrightarrow (M_1+M_2)/M_1, \qquad x \longmapsto x+M_1.$$

It is onto because $(u+x)+M_1 = x+M_1$ for $u \in M_1, x \in M_2$, and

$$\ker\beta = \{x \in M_2 : x \in M_1\} = M_1 \cap M_2.$$

Both maps are linear. Apply the first isomorphism theorem to each. □

Example 3.14. Inside $\mathbb{Z}$, put $M_1 = 4\mathbb{Z}$ and $M_2 = 6\mathbb{Z}$. Their sum is $2\mathbb{Z}$ and their intersection is $12\mathbb{Z}$. The second identity becomes

$$2\mathbb{Z}/4\mathbb{Z} \cong 6\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}.$$

The map from right to left sends the class of 6 to its class modulo 4, which is the nonzero class $2+4\mathbb{Z}$. This is an isomorphism of modules; the quotient notation does not require any of these subgroups to be unital subrings.

For submodules $N, P \subseteq M$, define their *colon ideal*

$$(N : P) = \{a \in A : aP \subseteq N\}.$$

It is an ideal: if a, b have the required property, so do $a+b$ and ra, for every $r \in A$. This notation records the scalars that move the whole submodule P into N.

For two arbitrary modules, $M \oplus N$ denotes their *direct sum*: its elements are pairs (m,n), with coordinatewise addition and scalar multiplication. This construction does not require M, N to be submodules of a common module. It will be studied systematically in Lecture 4.

Proposition 3.15 (Annihilator identities; AM in-text Exercise 2.2)**.** *If M,N are submodules of a common module, then* $\operatorname{Ann}(M+N) = \operatorname{Ann}(M) \cap \operatorname{Ann}(N)$. *If N,P are submodules of a module, then* $(N : P) = \operatorname{Ann}((N+P)/N)$.

Proof. For any scalar $a \in A$,

$$\begin{aligned} a \in \operatorname{Ann}(M+N) &\iff a(m+n) = 0 \quad (m \in M, n \in N) \\ &\iff aM = 0 \text{ and } aN = 0 \\ &\iff a \in \operatorname{Ann}(M) \cap \operatorname{Ann}(N). \end{aligned}$$

The middle equivalence uses $n = 0$, $m = 0$, and $a(m+n) = am + an$. The same calculation gives the direct-sum formula for arbitrary modules. Finally,

$$\begin{aligned} a \in \operatorname{Ann}((N+P)/N) &\iff a(p+N) = 0 \quad (p \in P) \\ &\iff aP \subseteq N \\ &\iff a \in (N : P). \end{aligned}$$

□

Example 3.16. In $\mathbb{Z}$, $(6\mathbb{Z} : 4\mathbb{Z}) = (3)$, because $6 \mid 4a$ exactly when $3 \mid a$. The quotient $(6\mathbb{Z} + 4\mathbb{Z})/6\mathbb{Z} = 2\mathbb{Z}/6\mathbb{Z}$ has three elements, so its annihilator is also (3). The two sides of the formula answer the same question in different forms.

Solving for all homomorphisms between two cyclic modules

Let r,s be positive integers and $d = \gcd(r,s)$. A homomorphism $\mathbb{Z}/r\mathbb{Z} \to \mathbb{Z}/s\mathbb{Z}$ is determined by the image u of $\bar{1}$, subject to $ru = 0$ in $\mathbb{Z}/s\mathbb{Z}$. Write $r = dr'$, $s = ds'$ with $\gcd(r',s') = 1$. For an integer representative of u,

$$ru \equiv 0 \pmod{s} \iff s' \mid r'u \iff s' \mid u.$$

The last implication follows from Bézout's identity for r',s'. The possibilities are

$$u = 0, s', 2s', \ldots, (d-1)s' \pmod{s}.$$

Addition of maps corresponds to addition of these residues. Consequently

$$\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/r\mathbb{Z}, \mathbb{Z}/s\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z}.$$

For $r = 6$, $s = 15$, the three maps send $\bar{1}$ to $0, 5, 10$. A choice such as $\bar{1} \mapsto 1$ is not allowed: it does not respect the defining relation $6\bar{1} = 0$. In a presentation, respecting the relations is precisely the condition that a proposed assignment of generators define a homomorphism.

3.7 Exercises

Exercise 3.1 (Annihilators and quotient scalars)**.** Let A be a ring, $I \subseteq A$ an ideal, and M an A-module. Prove that $\operatorname{Ann}_A(A/I) = I$. Show that if $I \subseteq \operatorname{Ann}_A(M)$, then M has a well-defined A/I-module structure.

Exercise 3.2 (Simplifying a presentation)**.** Let M be generated over $\mathbb{Z}$ by u, v, with relations $2u = 0$ and $3v = u$. Identify M.

Exercise 3.3 (A nilpotent operator as a module)**.** Let $V = k^2$ have basis e_1, e_2, and let $T(e_1) = 0$, $T(e_2) = e_1$. Make V a $k[t]$-module by letting t act as T. Show that this module is isomorphic to $k[t]/(t^2)$.

Exercise 3.4 (A Hom calculation)**.** Compute the abelian group

$$\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/6\mathbb{Z}).$$

Then explain why $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/4\mathbb{Z}, \mathbb{Z}) = 0$.

Lecture 4
Direct sums and exact sequences

Direct sums combine modules, while exact sequences record how a submodule sits inside a module and how the quotient is formed. The splitting criterion identifies exactly when this construction gives a direct sum. Reading: [1, Chapter 2, sections on direct sum and product, exact sequences, and Hom].

4.1 Direct sums and direct products

For two modules M, N, their *direct sum* $M \oplus N$ is the set of pairs (m,n), with coordinatewise addition and scalar multiplication. The following maps are *canonical*: their formulas require no choice of basis or representatives. They are

$$\iota_M(m) = (m,0), \quad \iota_N(n) = (0,n), \quad p_M(m,n) = m, \quad p_N(m,n) = n.$$

Every element has the unique expression $(m,n) = \iota_M(m) + \iota_N(n)$. Thus a map $M \oplus N \to P$ is determined by its restrictions to the two summands, while a map $P \to M \oplus N$ is determined by its two coordinate maps.

For a family $(M_i)_{i \in S}$, define

$$\prod_{i \in S} M_i = \{(m_i) : m_i \in M_i \text{ for every } i\},$$

$$\bigoplus_{i \in S} M_i = \{(m_i) \in \prod_i M_i : m_i = 0 \text{ for all but finitely many } i\}.$$

The *support of a tuple* (m_i) is the set of indices i for which $m_i \neq 0$; the direct sum consists of tuples with finite support. This use of the word support is distinct from the prime-ideal support introduced in Lecture 8. The sum and product coincide when S is finite. For infinite S, the finite-support condition is essential. The sequence $(1,1,1,\ldots)$ belongs to $\prod_{n \geq 1} \mathbb{Z}$ but not to $\bigoplus_{n \geq 1} \mathbb{Z}$.

An identification of sets of homomorphisms is called *natural* when it is compatible with composition with homomorphisms in the variables. For example, a map $P \to Q$ should

give the same result whether it is composed before or after the identification. A *universal property* specifies an object by maps into or out of it; the following proposition gives two such properties.

Proposition 4.1 (Mapping properties of sums and products)**.** *For every module P, there are natural bijections*

$$\operatorname{Hom}_A\left(\bigoplus_i M_i, P\right) \simeq \prod_i \operatorname{Hom}_A(M_i, P),$$

$$\operatorname{Hom}_A\left(P, \prod_i M_i\right) \simeq \prod_i \operatorname{Hom}_A(P, M_i).$$

Here "natural" means the identifications commute with composing the maps with other module homomorphisms.

Proof. Given maps $f_i : M_i \to P$, define $f((m_i)) = \sum_i f_i(m_i)$. This sum is finite because (m_i) lies in the direct sum. It defines a linear map, and any map with the required restrictions must use this formula. Conversely, restricting f to the ith summand recovers f_i.

For the second assertion, given maps $g_i : P \to M_i$, define $g(p) = (g_i(p))_i$. No finite-support requirement is imposed on a product, so this tuple always belongs to the target. Coordinate projections recover the g_i, proving the inverse correspondence. Composing any of these formulas with another linear map verifies the stated naturality. □

The formulas in the proof show how a universal property constructs a map from simpler component maps. The same method will construct maps from tensor products.

For submodules $U, V \subseteq M$, the addition map

$$\alpha : U \oplus V \to M, \qquad (u, v) \mapsto u + v$$

satisfies

$$\operatorname{im}\alpha = U + V, \qquad \ker\alpha = \{(u, -u) : u \in U \cap V\}.$$

Thus M is the *internal direct sum* of U, V precisely when

$$U + V = M, \quad U \cap V = 0 \quad \Longleftrightarrow \quad \alpha \text{ is an isomorphism.}$$

Equivalently, each $m \in M$ has a unique expression $m = u + v$.

4.2 Exact sequences

Recall Definition 3.8: a sequence is exact at a module when the image of the incoming map equals the kernel of the outgoing map. In particular, a short exact sequence has the form

$$0 \longrightarrow L \xrightarrow{u} M \xrightarrow{v} N \longrightarrow 0.$$

The three exactness conditions are

$$\ker u = 0, \qquad \ker v = u(L), \qquad \operatorname{im} v = N.$$

Consequently the first isomorphism theorem gives

$$M/u(L) \xrightarrow{\sim} N, \qquad m + u(L) \longmapsto v(m).$$

A short exact sequence specifies a submodule and its quotient together with their maps.

The equation $vu = 0$ follows from exactness, but is weaker than exactness: it only gives $\operatorname{im} u \subseteq \ker v$. For example, the zero maps $k \xrightarrow{0} k \xrightarrow{0} k$ compose to zero, but the image of the first is 0 and the kernel of the second is k.

Example 4.2 (A presentation revisited)**.** The sequence

$$A^r \xrightarrow{D} A^n \xrightarrow{\pi} M \longrightarrow 0$$

is exact precisely when π is surjective and $\ker \pi = \operatorname{im} D$. These are the assertions that the chosen symbols generate M and the columns of D generate *all* relations. Merely verifying $\pi D = 0$ proves only that the proposed relations are valid; it does not prove they suffice.

4.3 Split exact sequences

For maps $L \xrightarrow{u} M \xrightarrow{v} N$, a *section* of v is a linear map $s : N \to M$ with $vs = \mathrm{id}_N$, and a *retraction* of u is a linear map $r : M \to L$ with $ru = \mathrm{id}_L$. Thus a section chooses a preimage of every element in a way compatible with addition and scalars; a retraction returns each element of $u(L)$ to its original element of L. A short exact sequence is called *split* if it admits a section. The next theorem gives equivalent descriptions.

Theorem 4.3 (Splitting criterion)**.** *For a short exact sequence*

$$0 \longrightarrow L \xrightarrow{u} M \xrightarrow{v} N \longrightarrow 0,$$

the following are equivalent:

1. *There is an A-linear map $s : N \to M$ with $vs = \mathrm{id}_N$; it is called a* section *of v.*
2. *There is an A-linear map $r : M \to L$ with $ru = \mathrm{id}_L$; it is called a* retraction *of u.*
3. *There is an isomorphism $L \oplus N \simeq M$ carrying $(l,0)$ to $u(l)$ and whose composite with v is $(l,n) \mapsto n$.*

When these conditions hold, the sequence is split exact*.*

Proof. Suppose a section s exists. Define $\Phi : L \oplus N \to M$ by $\Phi(l,n) = u(l) + s(n)$. Given $m \in M$, put $n = v(m)$. Then

$$v(m - s(n)) = v(m) - vs(n) = n - n = 0,$$

so $m - s(n) = u(l)$ for a unique l, by exactness and injectivity of u. This proves surjectivity of Φ. If $u(l) + s(n) = 0$, applying v gives $n = 0$, and then injectivity gives $l = 0$. Thus Φ is an isomorphism with the required compatibility. Projection onto L through Φ^{-1} gives a retraction.

Conversely, suppose a retraction r exists. For $n \in N$, choose a lift $m \in M$ with $v(m) = n$, and set $s(n) = m - u(r(m))$. If another lift is $m + u(l)$, then

$$(m + u(l)) - u(r(m + u(l))) = m + u(l) - u(r(m) + l) = m - u(r(m)).$$

Therefore s is well defined. Choosing sums and scalar multiples of lifts proves that it is linear, and $vs(n) = n$. Finally, an isomorphism in condition (3) directly yields a section by $n \mapsto \Phi(0, n)$. This proves all equivalences. □

Corollary 4.4. *A short exact sequence whose quotient N is a free module splits. In particular, every short exact sequence of vector spaces splits.*

Proof. Choose a basis (e_i) of N and lift each e_i to an element $m_i \in M$. There is a unique linear map $s : N \to M$ sending e_i to m_i, because every vector has a unique finite basis expansion. Then $vs(e_i) = e_i$ for each basis element, hence $vs = \mathrm{id}_N$. For vector spaces, a basis exists, so the same construction applies. The chosen lifts explain why the splitting is generally not canonical. □

Example 4.5 (An exact sequence that does not split)**.** Consider

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{u} \mathbb{Z}/4\mathbb{Z} \xrightarrow{v} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0, \qquad u(\bar{1}) = \bar{2},$$

where v is reduction modulo 2. The image of u is $\{\bar{0}, \bar{2}\} = \ker v$, so the sequence is exact. If a section existed, $s(\bar{1})$ would be odd, hence either $\bar{1}$ or $\bar{3}$. But both have order 4, whereas linearity requires $2s(\bar{1}) = s(2\bar{1}) = 0$. This contradiction proves nonsplitting. Equivalently, $\mathbb{Z}/4\mathbb{Z}$ cannot be isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$: the former has an element of order 4, and the latter does not.

The obstruction to splitting an extension of cyclic groups

Consider

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{i} \mathbb{Z}/4\mathbb{Z} \xrightarrow{q} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0, \qquad i(\bar{a}) = \overline{2a}, \quad q(\bar{b}) = \bar{b}.$$

The map i is well defined because changing a by 2 changes $2a$ by 4. Its kernel is zero, and its image is $\{0, 2\}$, which is exactly the kernel of q. This verifies the sequence before asking whether it splits.

A section s of q would have to send $\bar{1}$ to 1 or 3 modulo 4. Since s is a homomorphism, it must also satisfy $2s(\bar{1}) = s(0) = 0$. But twice either possible lift is 2 modulo 4. No section exists. Equivalently, $\mathbb{Z}/4\mathbb{Z}$ has an element of order four, whereas every element of $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ is killed by two.

More generally, projectivity has three equivalent descriptions for a module P:

(i) every map $P \to N$ lifts through every surjection $M \to N$;

(ii) every short exact sequence ending in P splits;

(iii) P is a direct summand of a free module.

For (i)⇒(ii), lift id_P. For (ii)⇒(iii), choose a free module F with a surjection $F \to P$; splitting identifies F with $\ker(F \to P) \oplus P$. For (iii)⇒(i), write $F = P \oplus Q$. Compose a proposed map $P \to N$ with the projection $F \to P$, lift the resulting map using a basis of F, and restrict the lift to P. These constructions work for an infinite basis as well. A projective module need not have a basis itself.

4.4 Exactness after taking Hom

Fix an A-module P. A map $u : L \to M$ gives a map

$$u_* : \mathrm{Hom}_A(P,L) \longrightarrow \mathrm{Hom}_A(P,M), \qquad f \longmapsto u \circ f.$$

This is called *postcomposition*. If $0 \to L \xrightarrow{u} M \xrightarrow{v} N \to 0$ is exact, then

$$0 \longrightarrow \mathrm{Hom}_A(P,L) \xrightarrow{u_*} \mathrm{Hom}_A(P,M) \xrightarrow{v_*} \mathrm{Hom}_A(P,N)$$

is exact. This property is called *left exactness* of $\mathrm{Hom}_A(P,-)$.

Indeed, injectivity of u gives

$$u \circ f = 0 \quad \Longrightarrow \quad f = 0,$$

so u_* is injective. If $v \circ g = 0$, then

$$g(P) \subseteq \ker v = \mathrm{im}\, u.$$

For each p, define $f(p)$ as the unique element satisfying $u(f(p)) = g(p)$. For $p,q \in P$ and $a,b \in A$,

$$u(f(ap+bq)) = g(ap+bq) = u(af(p)+bf(q)).$$

Cancel u to prove linearity. Thus $g = u_* f$, proving $\ker v_* \subseteq \mathrm{im}\, u_*$. The reverse inclusion follows from $v_* u_* f = vuf = 0$.

There need not be a zero on the right. Surjectivity there would say that every map $P \to N$ can be lifted to M, which is an additional property.

Precomposition gives a second version, in which the direction of the arrows is reversed. Given a short exact sequence as above and a fixed target P,

$$0 \longrightarrow \mathrm{Hom}_A(N,P) \xrightarrow{v^*} \mathrm{Hom}_A(M,P) \xrightarrow{u^*} \mathrm{Hom}_A(L,P)$$

is exact, where $v^*(h) = h \circ v$ and $u^*(g) = g \circ u$. The arrows reverse because we are *precomposing*. To prove exactness, a map $M \to P$ that vanishes on $u(L)$ factors uniquely through

$M/u(L) \simeq N$. Injectivity on the left follows from surjectivity of v. Again, extending an arbitrary map from L to M is not automatic.

Remark 4.6 (A first look at projectivity). A module P is called *projective* if every map $P \to N$ lifts through every surjection $M \to N$. Thus projectivity is precisely the missing surjectivity condition in the first Hom sequence. Every free module is projective: choose lifts of the images of a basis and extend linearly. In particular, the lifting property of free modules is the basis-lifting construction already used in the splitting criterion.

4.5 Exactness criteria using Hom

The left-exactness calculations above are half of a useful equivalence. In the following statements, "for every module" matters: a single poorly chosen test module may detect nothing. The result is [1, Proposition 2.9].

Proposition 4.7 (Both exactness tests; AM 2.9)**.**

1. *A sequence* $L \xrightarrow{u} M \xrightarrow{v} N \to 0$ *is exact if and only if, for every* A*-module* P*,*

$$0 \to \mathrm{Hom}_A(N,P) \xrightarrow{v^*} \mathrm{Hom}_A(M,P) \xrightarrow{u^*} \mathrm{Hom}_A(L,P)$$

   is exact.
2. *A sequence* $0 \to L \xrightarrow{u} M \xrightarrow{v} N$ *is exact if and only if, for every* A*-module* P*,*

$$0 \to \mathrm{Hom}_A(P,L) \xrightarrow{u_*} \mathrm{Hom}_A(P,M) \xrightarrow{v_*} \mathrm{Hom}_A(P,N)$$

   is exact.

Proof. For (1), suppose first that the original sequence is exact. A map from N is determined by its composite with the surjection v, proving injectivity of v^*. A map $h : M \to P$ satisfies $hu = 0$ exactly when it factors through the quotient $M/u(L) = N$. This proves exactness in the middle.

Conversely, suppose all displayed Hom sequences are exact. Let $P = \operatorname{coker} v$ and let $q : N \to P$ be the quotient map. Since $qv = 0$ and v^* is injective, $q = 0$. A quotient map is onto, so $P = 0$, and v is onto. Taking $P = N$ and the identity map on N, the assertion $u^*v^* = 0$ gives $vu = 0$. Hence $u(L) \subseteq \ker v$. Now take $P = M/u(L)$ and the quotient $\pi : M \to P$. Since $\pi u = 0$, exactness gives $h : N \to P$ with $\pi = hv$. If $v(m) = 0$, then $\pi(m) = h(0) = 0$, so $m \in u(L)$. This is the opposite inclusion.

For (2), the forward implication follows from the element-by-element lifting into $\ker v = u(L)$ proved above; no surjectivity of v is needed. Conversely, put $P = \ker u$. Its inclusion $j : P \to L$ satisfies $uj = 0$. Injectivity of u_* forces $j = 0$, so $\ker u = 0$. Taking $P = L$ and id_L, the equality $v_*u_* = 0$ gives $vu = 0$. Finally put $P = \ker v$ and consider its inclusion $i : P \to M$. Exactness of Hom supplies $h : P \to L$ with $i = uh$. Thus every element of $\ker v$ lies in $u(L)$; the reverse inclusion follows from $vu = 0$. □

4.6 The snake lemma

A diagram of maps is *commutative* if composing along any two directed paths with the same starting and ending objects gives the same map. For the diagram below, this means $bu = u'a$ and $cv = v'b$. Suppose two short exact sequences are related by three such compatible maps. The snake lemma says how the failure of injectivity and surjectivity in those three maps is related. Its name comes from drawing the boundary map across the diagram; the proof is an ordinary calculation with representatives.

Proposition 4.8 (Snake lemma; AM 2.10)**.** *Suppose the following diagram commutes and both rows are exact:*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & L & \xrightarrow{u} & M & \xrightarrow{v} & N & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle a} & & \downarrow{\scriptstyle b} & & \downarrow{\scriptstyle c} & & \\ 0 & \longrightarrow & L' & \xrightarrow{u'} & M' & \xrightarrow{v'} & N' & \longrightarrow & 0. \end{array}$$

There is an exact sequence

$$\begin{aligned} 0 \longrightarrow \ker a \longrightarrow \ker b \longrightarrow \ker c \xrightarrow{\delta} \operatorname{coker} a \\ \longrightarrow \operatorname{coker} b \longrightarrow \operatorname{coker} c \longrightarrow 0. \end{aligned}$$

The first two maps are restrictions of u, v, and the last two are induced by u', v'.

Proof. **Constructing the boundary.** Take $n \in \ker c$. Choose $m \in M$ with $v(m) = n$, using surjectivity of v. Commutativity gives $v'b(m) = cv(m) = c(n) = 0$. Thus $b(m)$ lies in $\ker v' = u'(L')$. There is a unique $l' \in L'$ with $u'(l') = b(m)$, because u' is injective. Define $\delta(n) = l' + a(L) \in \operatorname{coker} a$.

Independence and linearity. If $\widetilde{m}$ is another lift, then $\widetilde{m} - m = u(l)$ for some $l \in L$. Consequently $b(\widetilde{m}) - b(m) = bu(l) = u'a(l)$. The corresponding elements of L' differ by $a(l)$, so their cokernel classes agree. To compute $\delta(n_1 + n_2)$, use the sum of chosen lifts; to compute $\delta(rn)$, use r times a lift. These choices prove additivity and A-linearity.

At $\ker a$**.** Restriction of u is injective because u is injective on all of L.

At $\ker b$**.** If $m \in \ker b$ and $v(m) = 0$, write $m = u(l)$. Then $0 = b(m) = u'a(l)$, and injectivity of u' gives $a(l) = 0$. Thus m comes from $\ker a$. Conversely $vu = 0$, and $bu(l) = u'a(l) = 0$ for $l \in \ker a$.

At $\ker c$**.** If $n = v(m)$ for $m \in \ker b$, the definition gives $\delta(n) = 0$. Conversely, suppose $\delta(n) = 0$. For a chosen lift m we have $b(m) = u'(l')$ with $l' = a(l)$ for some $l \in L$. Replace m by $m - u(l)$. It still maps to n, and $b(m - u(l)) = u'a(l) - u'a(l) = 0$. Thus n comes from $\ker b$.

At $\operatorname{coker} a$**.** The induced map sends $[l']$ to $[u'(l')]$. If $[l'] = \delta(n)$, the construction gives $u'(l') = b(m)$, so its image is zero. Conversely, if $[u'(l')] = 0$, choose m with $b(m) = u'(l')$. Then $c(v(m)) = v'b(m) = v'u'(l') = 0$, and the definition gives $\delta(v(m)) = [l']$.

At $\operatorname{coker} b$**.** Let $[m'] \in \operatorname{coker} b$ have zero image in $\operatorname{coker} c$. Then $v'(m') = c(n)$ for some $n \in N$. Lift n to $m \in M$. The difference $m' - b(m)$ is killed by v', so it equals $u'(l')$ for

some l'. Hence $[m'] = [u'(l')]$, as required. Conversely $v'u' = 0$, so every such class has zero image.

At $\operatorname{coker} c$**.** Every class has a representative $n' \in N'$. Surjectivity of v' gives m' with $v'(m') = n'$, and its class maps to the desired class. This proves all the exactness assertions. □

Example 4.9 (The boundary remembers a non-split extension)**.** Take both rows to be $0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$, and take each vertical map to be multiplication by 2. The vertical map on $\mathbb{Z}/2\mathbb{Z}$ is zero. The nonzero class $\bar{1} \in \ker c$ lifts to $1 \in \mathbb{Z}$. Its image under the middle vertical map is $2 = u'(1)$, so

$$\delta : \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z}, \qquad \bar{1} \longmapsto \bar{1}.$$

Thus the boundary is an isomorphism. A map built from kernels and cokernels can carry real information even when the two vertical maps on free modules are injective.

4.7 Additivity and finite generation in exact sequences

Proposition 4.10 (Additive functions on an exact sequence; AM 2.11)**.** *Let λ assign elements of an abelian group G to a class $\mathscr{C}$ of modules, and suppose $\lambda(M) = \lambda(L) + \lambda(N)$ for every short exact sequence $0 \to L \to M \to N \to 0$ whose terms lie in $\mathscr{C}$. If $0 \to M_0 \to M_1 \to \cdots \to M_n \to 0$ is exact and all its modules and kernels belong to $\mathscr{C}$, then*

$$\sum_{i=0}^{n} (-1)^i \lambda(M_i) = 0.$$

Proof. Put $K_i = \ker(M_i \to M_{i+1})$, with $K_0 = 0$ and $K_{n+1} = 0$. Exactness gives short exact sequences $0 \to K_i \to M_i \to K_{i+1} \to 0$. The hypothesis therefore gives $\lambda(M_i) = \lambda(K_i) + \lambda(K_{i+1})$. In the alternating sum, each intermediate $\lambda(K_i)$ occurs twice with opposite signs. The endpoint terms vanish: additivity applied to the zero sequence gives $\lambda(0) = 0$. Only zero remains. □

For vector spaces this says that the alternating sum of dimensions in a finite exact sequence is zero. For example, $0 \to k \to k^3 \to k^2 \to 0$ gives $1 - 3 + 2 = 0$. Dimension alone does not prove exactness: it is a consequence of the actual kernel-and-image statements.

Two failures of right exactness for Hom

Apply $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, -)$ to the surjection $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. Any map $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$ is zero because its image is killed by two and $\mathbb{Z}$ has no nonzero element killed by two. Therefore

$$\mathrm{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = 0, \qquad \mathrm{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z},$$

and the induced map is not onto. The identity of $\mathbb{Z}/2\mathbb{Z}$ has no lift to $\mathbb{Z}$.

There is also a failure for the contravariant variable. In the inclusion $2\mathbb{Z} \hookrightarrow \mathbb{Z}$, the homomorphism $h : 2\mathbb{Z} \to \mathbb{Z}$ given by $h(2n) = n$ cannot extend to $\mathbb{Z}$. Such an extension g would require $2g(1) = 1$. Under the identifications $\mathrm{Hom}(\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}$ and $\mathrm{Hom}(2\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}$ obtained by evaluating at 1 and 2, restriction is multiplication by two. Its cokernel is $\mathbb{Z}/2\mathbb{Z}$. In the first example the missing map is a lift; in the second it is an extension from a submodule. Keeping the two variables of Hom distinct prevents an incorrect reversal of arrows.

4.8 Exercises

Exercise 4.1 (Why the infinite sum cannot be ignored)**.** Let $M = \bigoplus_{n \geq 1} \mathbb{Z}$. Show that $s((a_n)) = \sum_n a_n$ defines a homomorphism $M \to \mathbb{Z}$. Explain why the same formula does not define a map on $\prod_{n \geq 1} \mathbb{Z}$.

Exercise 4.2 (Check every position)**.** For an integer $n \geq 2$, verify that

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\times n} \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow 0$$

is exact. Why does the analogous sequence beginning with multiplication by $\bar{2}$ on $\mathbb{Z}/6\mathbb{Z}$ fail to be exact at its first nonzero term?

Exercise 4.3 (All splittings of a simple sequence)**.** Let A be a ring. Define $v : A^2 \to A$ by $v(a,b) = a+b$ and $u : A \to A^2$ by $u(t) = (t,-t)$. Verify that $0 \to A \xrightarrow{u} A^2 \xrightarrow{v} A \to 0$ is exact, and give two distinct sections of v when $A \neq 0$.

Exercise 4.4 (Hom need not preserve surjections)**.** Apply $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, -)$ to the surjection $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. Is the resulting map surjective?

Exercise 4.5 (Finite generation in an extension; AM Chapter 2, Exercise 9)**.** In a short exact sequence $0 \to L \xrightarrow{u} M \xrightarrow{v} N \to 0$ of A-modules, assume that L and N are finitely generated over A. Prove that M is finitely generated.

Lecture 5
Tensor products: making bilinear maps linear

A tensor product replaces bilinear maps by linear maps. Its construction by generators and relations provides both explicit calculations and a proof of its behavior on exact sequences. Reading: [1, Chapter 2, the sections on tensor products and exactness properties of the tensor product].

5.1 Bilinear maps and the tensor product

An *A-bilinear map* $b : M \times N \to P$ is linear in each variable separately:

$$b(m+m',n) = b(m,n) + b(m',n), \qquad b(m,n+n') = b(m,n) + b(m,n'),$$

$$b(am,n) = a\,b(m,n) = b(m,an).$$

For example, multiplication $A \times A \to A$ is bilinear. The map $(u,v) \mapsto uv^{\mathsf{T}}$ from pairs of column vectors in k^r, k^s to $r \times s$ matrices is also bilinear; here v^{T} is the row vector obtained by transposing v, so the (i,j) entry of the matrix is $u_i v_j$.

Bilinearity is different from linearity on the direct sum. A linear map $M \oplus N \to P$ has the form $f(m) + g(n)$; a bilinear map vanishes whenever either argument is zero and records an interaction between both arguments. We seek a module whose ordinary linear maps correspond exactly to bilinear maps out of $M \times N$.

Definition 5.1 (Tensor product and universal property)**.** A tensor product of M and N over A is an A-module $M \otimes_A N$ with a bilinear map

$$\tau : M \times N \to M \otimes_A N, \qquad (m,n) \mapsto m \otimes n,$$

such that for every bilinear map $b : M \times N \to P$, there is a unique linear map $\widetilde{b} : M \otimes_A N \to P$ satisfying $\widetilde{b}(m \otimes n) = b(m,n)$.

The definition is a requirement, so we must prove that such an object exists. The following construction also explains exactly which relations the tensor symbols obey.

5.2 Construction by generators and relations

Let F be the free A-module with one formal basis symbol $e_{m,n}$ for each pair $(m,n) \in M \times N$. An element of F is a finite sum of scalar multiples of these symbols. Let R be the submodule generated by

$$e_{m+m',n} - e_{m,n} - e_{m',n}, \qquad e_{m,n+n'} - e_{m,n} - e_{m,n'},$$
$$e_{am,n} - ae_{m,n}, \qquad e_{m,an} - ae_{m,n},$$

for all permissible elements and scalars. Define

$$M \otimes_A N = F/R, \qquad m \otimes n = e_{m,n} + R.$$

The defining relations make τ bilinear. They also imply

$$(am) \otimes n = m \otimes (an),$$

usually called the *balancing relation*. It allows an A-scalar to move from one factor to the other.

Theorem 5.2 (Existence and uniqueness of tensor products; AM 2.12)**.** *For any two A-modules M,N, there exist an A-module T and an A-bilinear map $\tau : M \times N \to T$ such that every A-bilinear map $b : M \times N \to P$ factors uniquely as $b = \widetilde{b} \circ \tau$, with $\widetilde{b} : T \to P$ linear. The construction above gives such a pair. Any two such pairs are uniquely isomorphic by an isomorphism preserving their canonical bilinear maps.*

Proof. Given b, define

$$\beta : F \longrightarrow P, \qquad \sum_i a_i e_{m_i,n_i} \longmapsto \sum_i a_i b(m_i,n_i).$$

Bilinearity gives, for example,

$$\beta(e_{m+m',n} - e_{m,n} - e_{m',n}) = b(m+m',n) - b(m,n) - b(m',n) = 0,$$
$$\beta(e_{am,n} - ae_{m,n}) = b(am,n) - ab(m,n) = 0.$$

The two relations in the second variable are identical calculations. Hence $R \subseteq \ker \beta$, and the quotient factorization property yields

$$\widetilde{b} : F/R \to P, \qquad m \otimes n \mapsto b(m,n).$$

Pure tensors generate F/R, so this map is unique.

For two pairs $(T,\tau),(T',\tau')$ with the required property, obtain unique maps $\alpha : T \to T'$ and $\gamma : T' \to T$ with

$$\alpha\tau = \tau', \qquad \gamma\tau' = \tau.$$

Then $(\gamma\alpha)\tau = \tau$ and $(\alpha\gamma)\tau' = \tau'$. Uniqueness gives $\gamma\alpha = \mathrm{id}_T$ and $\alpha\gamma = \mathrm{id}_{T'}$. □

Every tensor is a finite sum $\sum_i m_i \otimes n_i$: a scalar coefficient can be absorbed into either factor. An individual $m \otimes n$ is called a *pure tensor*. Do not assume every tensor is pure, or that $m \otimes n = 0$ forces one of its factors to be zero.

For a matrix over a field, its *rank* is the dimension of the image of its associated linear map. The matrix unit E_{ij} has entry 1 in position (i, j) and 0 in every other position.

Example 5.3 (A tensor that is not pure)**.** Let $V = W = k^2$, with standard bases e_1, e_2 and f_1, f_2. The bilinear matrix map induces a linear map $V \otimes_k W \to \mathrm{Mat}_{2\times 2}(k)$ sending $e_i \otimes f_j$ to the matrix unit E_{ij}. Conversely, the map sending E_{ij} to $e_i \otimes f_j$ is an inverse, as one checks by expanding vectors in their bases. Under this isomorphism,

$$e_1 \otimes f_1 + e_2 \otimes f_2$$

corresponds to the identity matrix. A pure tensor corresponds to uv^{T}, whose image is contained in the one-dimensional span of u, so it has rank at most one. The identity matrix has rank two. Hence this tensor is not pure.

5.3 Induced maps and basic tensor identities

If $f : M \to M'$ and $g : N \to N'$ are linear, there is an induced map

$$f \otimes g : M \otimes_A N \longrightarrow M' \otimes_A N', \qquad m \otimes n \longmapsto f(m) \otimes g(n).$$

The universal property proves that this is well defined. Checking its values on pure tensors proves the composition rule $(f' \otimes g')(f \otimes g) = (f'f) \otimes (g'g)$.

Proposition 5.4. *There are canonical isomorphisms*

$$A \otimes_A M \simeq M, \qquad M \otimes_A N \simeq N \otimes_A M,$$

$$\left(\bigoplus_i M_i \right) \otimes_A N \simeq \bigoplus_i (M_i \otimes_A N).$$

In particular, $A^r \otimes_A M \simeq M^r$, and if V, W are finite-dimensional vector spaces over k, then

$$\dim_k(V \otimes_k W) = (\dim_k V)(\dim_k W).$$

Proof. The bilinear map $(a, m) \mapsto am$ induces $A \otimes_A M \to M$. Its inverse is $m \mapsto 1 \otimes m$, because $a \otimes m = 1 \otimes am$. Swapping the variables gives $m \otimes n \mapsto n \otimes m$; swapping twice is the identity.

For the direct-sum assertion, send $(m_i)_i \otimes n$ to $(m_i \otimes n)_i$. Only finitely many m_i are nonzero, so this lies in the direct sum. The formula is bilinear and hence defines a linear map. In the reverse direction, use the inclusions $M_i \to \bigoplus_i M_i$ to send the ith component tensor $m_i \otimes n$ to its corresponding tensor in the larger module, and sum. These maps are inverse on pure tensors and on the generating component tensors, hence everywhere. Apply

this to copies of A for A^r. Finally, choosing bases identifies V and W with k^r, k^s, and the tensors of basis pairs form a basis indexed by rs pairs. □

5.4 Tensoring a quotient

Recall that IM is the submodule of finite sums $\sum_i a_i m_i$ with $a_i \in I$ and $m_i \in M$.

Theorem 5.5 (Quotient formula; AM Chapter 2, Exercise 2)**.** *There is a canonical isomorphism*

$$(A/I) \otimes_A M \simeq M/IM, \qquad (a+I) \otimes m \longmapsto am + IM.$$

Consequently,

$$(A/I) \otimes_A (A/J) \simeq A/(I+J).$$

Proof. Define maps in both directions by

$$\begin{aligned} \alpha((a+I) \otimes m) &= am + IM, \\ \beta(m + IM) &= (1+I) \otimes m. \end{aligned}$$

For $i \in I$, $(a+i)m - am = im \in IM$; thus the first formula is independent of a and is bilinear. The second descends from M because

$$(1+I) \otimes \sum_i a_i m_i = \sum_i (a_i + I) \otimes m_i = 0 \qquad (a_i \in I).$$

The composite calculations are

$$\begin{aligned} \alpha\beta(m + IM) &= m + IM, \\ \beta\alpha((a+I) \otimes m) &= (1+I) \otimes am = (a+I) \otimes m. \end{aligned}$$

Finally, for $M = A/J$,

$$I(A/J) = (I+J)/J, \qquad (A/J)/I(A/J) \cong A/(I+J).$$

□

Quotient tensors computed in both directions

The formula $(A/I)\otimes_A N \cong N/IN$ is a method for finding actual maps. Define

$$\alpha : (A/I)\otimes_A N \to N/IN, \quad (a+I)\otimes n \longmapsto an+IN,$$

and $\beta : N/IN \to (A/I)\otimes_A N$ by $n+IN \mapsto (1+I)\otimes n$. Changing a by an element of I changes an by an element of IN, so the first rule is well defined and balanced. To check the second, for $b \in I$ one has

$$(1+I)\otimes bn = (b+I)\otimes n = 0.$$

It therefore kills every finite sum defining IN. Finally

$$\alpha\beta(n+IN) = n+IN, \quad \beta\alpha((a+I)\otimes n) = (1+I)\otimes an = (a+I)\otimes n.$$

These checks show why the isomorphism is canonical.

Taking $N = A/J$ gives

$$(A/I)\otimes_A (A/J) \cong A/(I+J).$$

For integers $r,s>0$ this becomes $\mathbb{Z}/r\mathbb{Z}\otimes_{\mathbb{Z}}\mathbb{Z}/s\mathbb{Z} \cong \mathbb{Z}/\gcd(r,s)\mathbb{Z}$. In particular, $\mathbb{Z}/4\mathbb{Z}\otimes \mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$. The tensor $\bar{1}\otimes\bar{1}$ is killed by both four and six, hence by two; the displayed isomorphism also proves that this tensor is not zero.

5.5 Right exactness of tensor products

Theorem 5.6 (Tensor products preserve cokernels; AM 2.18)**.** *If*

$$L \xrightarrow{u} M \xrightarrow{v} N \longrightarrow 0$$

is exact, then for every A-module P,

$$L\otimes_A P \xrightarrow{u\otimes \mathrm{id}} M\otimes_A P \xrightarrow{v\otimes \mathrm{id}} N\otimes_A P \longrightarrow 0$$

is exact. This is called right exactness *of tensor product.*

Proof. Surjectivity on the right is concrete. Every element of $N\otimes_A P$ is a finite sum $\sum_i n_i\otimes p_i$. Choose lifts $m_i \in M$ with $v(m_i) = n_i$. Then $\sum_i m_i\otimes p_i$ maps to the chosen element.

To identify the kernel, put

$$Q = (M\otimes_A P)/\operatorname{im}(u\otimes \mathrm{id}).$$

Since $vu = 0$, the map $v\otimes\mathrm{id}$ descends to $\alpha : Q \to N\otimes_A P$. We construct its inverse. Given $n \in N$ and $p \in P$, choose a lift m of n and let

$$b(n,p) = \text{the class of } m\otimes p \text{ in } Q.$$

For another lift m', exactness gives $m' - m = u(l)$ for some $l \in L$. With brackets denoting classes in Q,

$$[m' \otimes p] - [m \otimes p] = [u(l) \otimes p] = [(u \otimes \mathrm{id})(l \otimes p)] = 0.$$

Thus b is well defined. Sums and scalar multiples of lifts prove bilinearity, so b induces $\beta : N \otimes_A P \to Q$. The composites satisfy

$$\alpha\beta(n \otimes p) = n \otimes p,$$
$$\beta\alpha([m \otimes p]) = [m \otimes p].$$

These generators determine the maps; hence α is an isomorphism and

$$\ker(v \otimes \mathrm{id}) = \mathrm{im}(u \otimes \mathrm{id}).$$

□

Example 5.7 (Computing from a presentation)**.** If $M = \mathrm{coker}(D : A^r \to A^n)$, right exactness gives

$$P^r \xrightarrow{D_P} P^n \longrightarrow M \otimes_A P \longrightarrow 0.$$

The matrix D_P uses the entries of D acting on P, so $M \otimes_A P \simeq \mathrm{coker}\, D_P$. For example,

$$M = \mathbb{Z}^2/\langle (2,0), (1,3) \rangle, \qquad D = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}.$$

For $P = \mathbf{F}_2$, reduction of the matrix gives

$$D_P = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \qquad M \otimes_{\mathbb{Z}} \mathbf{F}_2 \cong \mathbf{F}_2^2/\mathbf{F}_2(1,1) \cong \mathbf{F}_2.$$

The last isomorphism sends $(a,b) + \mathbf{F}_2(1,1)$ to $a+b$.

5.6 Flat modules and failure of left exactness

Tensor the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

with $\mathbb{Z}/2\mathbb{Z}$. Using the computations above, the part guaranteed to be exact becomes

$$\mathbb{Z}/2\mathbb{Z} \xrightarrow{0} \mathbb{Z}/2\mathbb{Z} \xrightarrow{\mathrm{id}} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

The first map is zero because multiplication by 2 acts as zero modulo 2. Thus it is not injective. One cannot place a zero on the left and still claim exactness. In particular, the leftmost zero of the original short exact sequence does not survive as an exactness assertion.

Definition 5.8. An A-module P is *flat* if tensoring with P preserves injections. Together with right exactness, this means that tensoring with P preserves all short exact sequences.

Every free module $F = A^{(S)} = \bigoplus_{s\in S} A$ is flat: if $u : L \to M$ is injective, the direct-sum tensor formula identifies

$$u \otimes \mathrm{id}_F : L \otimes_A F \longrightarrow M \otimes_A F \quad \text{with} \quad \bigoplus_{s\in S} u : \bigoplus_{s\in S} L \longrightarrow \bigoplus_{s\in S} M.$$

Its kernel is $\bigoplus_{s\in S} \ker u = 0$. Since every vector space has a basis, tensoring over a field is exact. The preceding example shows that $\mathbb{Z}/2\mathbb{Z}$ is not flat over $\mathbb{Z}$.

An explicit injection destroyed by tensoring

The inclusion $i : 2\mathbb{Z} \hookrightarrow \mathbb{Z}$ becomes the zero map after tensoring with $\mathbb{Z}/2\mathbb{Z}$. This should not be confused with the assertion that its source becomes zero. In fact

$$(2\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$$

because $2\mathbb{Z}$ is free on the element 2. Its nonzero generator is $2 \otimes \bar{1}$. The image in $\mathbb{Z} \otimes \mathbb{Z}/2\mathbb{Z}$ is

$$i(2) \otimes \bar{1} = 2 \otimes \bar{1} = 1 \otimes 2\bar{1} = 0.$$

The first tensor has its first factor in $2\mathbb{Z}$, where 1 is unavailable; the last calculation takes place after applying i. Writing the ambient tensor products is essential when comparing these expressions.

One general consequence is that a flat module over a domain is torsion-free. For $0 \neq a \in A$, multiplication $A \xrightarrow{a} A$ is injective. Tensoring with a flat module M identifies this map with $M \xrightarrow{a} M$, which must remain injective. The converse needs extra hypotheses on A and cannot be inferred from this argument.

5.7 Finite witnesses and several tensor factors

The tensor construction has a useful finiteness feature even when the modules themselves are very large.

Corollary 5.9 (A zero tensor has a finite witness; AM 2.13)**.** *If $x_i \in M$, $y_i \in N$, $1 \le i \le r$, and $\sum_i x_i \otimes y_i = 0$ in $M \otimes_A N$, there exist finitely generated submodules $M_0 \subseteq M$, $N_0 \subseteq N$ containing all the x_i, y_i, respectively, such that the same sum is zero in $M_0 \otimes_A N_0$.*

Proof. Return to the construction $M \otimes N = F/R$. The equality says that the formal sum $\sum_i e_{x_i,y_i}$ belongs to R. Membership in the submodule generated by the bilinearity relations means it is a *finite* linear combination of those relations. Collect every element of M

appearing as a first coordinate anywhere in this finite expression, including the original x_i, and let them generate M_0. Collect the second coordinates to generate N_0.

All the relations used in that expression are already valid relations for the construction of $M_0 \otimes N_0$. For example, a relation involving $x, x', x+x'$ is permitted because the chosen submodule contains the entries and is closed under addition. Thus the same finite calculation proves the equality in $M_0 \otimes N_0$. □

Remark 5.10. The map $M_0 \otimes N_0 \to M \otimes N$ induced by the inclusions of the factors need not be injective. The point is that suitable finite submodules can be chosen *after inspecting a witness for the equality*. An arbitrary finite submodule containing only the original displayed factors might not suffice.

A map $b : M_1 \times \cdots \times M_r \to P$ is *A-multilinear* if it is A-linear in any one variable when the other variables are fixed. Bilinear means the case $r = 2$. For example, multiplication $A^r \to A$, $(a_1, \ldots, a_r) \mapsto a_1 \cdots a_r$, is multilinear.

Proposition 5.11 (Several tensor factors; AM 2.12*)**.** *For A-modules $M_1, \ldots, M_r$, there exist an A-module T and an A-multilinear map $\tau : M_1 \times \cdots \times M_r \to T$ such that every multilinear map from this product to any A-module P factors uniquely through an A-linear map $T \to P$. The pair (T, τ) is unique up to a unique isomorphism preserving τ. We write $T = M_1 \otimes_A \cdots \otimes_A M_r$.*

Proof. Take the free module with a basis symbol $e_{m_1,\ldots,m_r}$ for each tuple. Quotient by the additivity and scalar-linearity relations in each individual position. For example, in position j impose

$$e_{\ldots,x+x',\ldots} - e_{\ldots,x,\ldots} - e_{\ldots,x',\ldots} = 0, \qquad e_{\ldots,ax,\ldots} - a e_{\ldots,x,\ldots} = 0.$$

The tuple-to-class map is multilinear by construction. A multilinear map to P first extends linearly from the basis to the free module. Each imposed relation maps to zero, so it descends to the quotient. Uniqueness follows because the tuple classes generate. If two pairs have this property, apply it to their canonical multilinear maps to obtain maps in both directions; each composite fixes the tuple classes and is therefore the identity. □

Proposition 5.12 (Canonical tensor identities, all parts; AM 2.14)**.** *For A-modules M, N, P, there are unique isomorphisms with the indicated formulas:*

1. *$M \otimes N \cong N \otimes M$, taking $m \otimes n$ to $n \otimes m$;*
2. *$(M \otimes N) \otimes P \cong M \otimes (N \otimes P) \cong M \otimes N \otimes P$, taking $(m \otimes n) \otimes p$ and $m \otimes (n \otimes p)$ to $m \otimes n \otimes p$;*
3. *$(M \oplus N) \otimes P \cong (M \otimes P) \oplus (N \otimes P)$, taking $(m,n) \otimes p$ to $(m \otimes p, n \otimes p)$;*
4. *$A \otimes M \cong M$, taking $a \otimes m$ to am.*

All tensor products in this proposition are over A.

Proof. For (1), the map $(m,n) \mapsto n \otimes m$ is bilinear and induces a linear map. The formula with the factors interchanged induces its inverse. Their composites fix pure tensors, hence all tensors.

For (2), put $T = M \otimes N \otimes P$. Fix $p \in P$; the map $(m,n) \mapsto m \otimes n \otimes p$ is bilinear, so it defines $h_p : M \otimes N \to T$. The dependence on p is linear, as one checks on the generators

$m \otimes n$. Thus $(t,p) \mapsto h_p(t)$ is bilinear and induces $(M \otimes N) \otimes P \to T$. In the other direction, $(m,n,p) \mapsto (m \otimes n) \otimes p$ is multilinear, so it gives $T \to (M \otimes N) \otimes P$. Their composites fix the generating tensors. The other parenthesization is treated by fixing m first: $(n,p) \mapsto m \otimes n \otimes p$ is bilinear, and the same two-step construction produces $M \otimes (N \otimes P) \cong T$.

For (3), the displayed formula is bilinear and gives a map f. The inclusions $M, N \to M \oplus N$ give maps $M \otimes P, N \otimes P \to (M \oplus N) \otimes P$. Adding them gives a map g from the direct sum. On a generating tensor,

$$gf((m,n) \otimes p) = (m,0) \otimes p + (0,n) \otimes p = (m,n) \otimes p.$$

The other composite fixes generators in each direct summand.

For (4), scalar multiplication is bilinear, so it gives a map to M. The inverse is $m \mapsto 1 \otimes m$, since $1 \otimes am = a \otimes m$. Uniqueness throughout follows from the specified values on generating tensors. This also explains the word "canonical": no basis choices enter the formulas. □

5.8 The finite test for flatness

Proposition 5.13 (Equivalent flatness criteria; AM 2.19)**.** *For an A-module P, the following are equivalent:*

1. *Tensoring with P preserves every exact sequence.*
2. *It preserves every short exact sequence.*
3. *For every injective map $f : L \to M$, the map $f \otimes \mathrm{id}_P$ is injective.*
4. *The preceding assertion holds whenever both L and M are finitely generated.*

Proof. Condition (1) immediately implies (2). Conversely split an exact sequence at its images into short exact sequences. Preservation of those sequences says that, at each tensor term, the outgoing map factors through the tensor of the next image by a surjection followed by an injection, and that its kernel is the preceding tensor image. This proves (1).

Right exactness, Proposition 5.6, shows that the only extra requirement in a short exact sequence is preservation of its first injection. Thus (2) and (3) are equivalent. Condition (3) plainly implies (4).

Assume (4), and let $f : L \to M$ be an arbitrary injection. Consider $z = \sum_{i=1}^{r} l_i \otimes p_i$ in its tensor kernel, so $\sum_i f(l_i) \otimes p_i = 0$ in $M \otimes P$. Let L_0 be generated by the l_i. Corollary 5.9 supplies a finitely generated $M_0 \subseteq M$ containing all $f(l_i)$, and a finitely generated submodule $P_0 \subseteq P$, such that the sum vanishes already in $M_0 \otimes P_0$. Its image in $M_0 \otimes P$ is therefore zero. Restriction gives an injection $f_0 : L_0 \to M_0$ between finitely generated modules. By (4), $f_0 \otimes \mathrm{id}_P$ is injective. Hence $\sum_i l_i \otimes p_i = 0$ in $L_0 \otimes P$, and mapping to $L \otimes P$ gives $z = 0$. This proves (3). □

Example 5.14. As an A-module, $A[x]$ is the direct sum of the copies $Ax^n \cong A$, for $n \geq 0$. Hence $A[x]$ is flat. This says that adding an unconstrained polynomial variable preserves all exact module calculations. Imposing a relation, as in $A/(a)$, can destroy this property.

5.9 Exercises

Exercise 5.1 (A tensor that vanishes)**.** Compute $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z}$ directly from the relations.

Exercise 5.2 (Integer and polynomial computations; AM Chapter 2, Exercise 1)**.** Compute $\mathbb{Z}/4\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/6\mathbb{Z}$ and $k[x,y]/(x) \otimes_{k[x,y]} k[x,y]/(y)$. Give the general formula for $\mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z}$, for positive m, n.

Exercise 5.3 (Do not confuse the submodule with its ambient image)**.** Let $I = 2\mathbb{Z} \subset \mathbb{Z}$. Compute $I \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z}$, and compute the map it induces to $\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z}$.

Exercise 5.4 (Tensor–Hom correspondence)**.** For A-modules M, N, P, construct a bijection

$$\operatorname{Hom}_A(M \otimes_A N, P) \simeq \operatorname{Hom}_A\big(M, \operatorname{Hom}_A(N, P)\big).$$

Exercise 5.5 (Direct sums and flatness; AM Chapter 2, Exercise 4)**.** For an arbitrary family $(P_i)_{i \in J}$ of A-modules, prove that $\bigoplus_{i \in J} P_i$ is flat over A if and only if every P_i is flat over A.

Exercise 5.6 (Tensor products and Hom)**.** Compute

$$\mathbb{Z}/6\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/10\mathbb{Z} \qquad \text{and} \qquad \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/6\mathbb{Z}, \mathbb{Z}/10\mathbb{Z}).$$

Then tensor the injection $\mathbb{Z} \xrightarrow{2} \mathbb{Z}$ with $\mathbb{Z}/2\mathbb{Z}$ and explain the result.

Lecture 6
Algebras and change of scalars

A module depends on its ring of scalars as well as on its underlying abelian group. Restriction and extension of scalars describe how these data change under a ring homomorphism. Applied to polynomial equations with a parameter, extension of scalars computes the equations at a chosen parameter value. Reading: [1, Chapter 2, sections on restriction and extension of scalars, algebras, and tensor products of algebras].

6.1 Algebras and their structure maps

Definition 6.1. An *A-algebra* is a ring B with a specified ring homomorphism $f : A \to B$, its *structure map*. If $g : A \to C$ is another structure map, an *A-algebra homomorphism* $h : B \to C$ is a ring homomorphism satisfying

$$h \circ f = g, \qquad h(f(a)) = g(a) \quad (a \in A).$$

The map f makes B an A-module by $a \cdot b = f(a)b$. This module structure is compatible with multiplication in B. The distinction between the ring and its structure map is important. The ring $k[x]$ can be made a $k[t]$-algebra by $t \mapsto x$, by $t \mapsto x^2$, or by $t \mapsto 0$; these maps define different scalar actions.

An A-algebra need not contain an isomorphic copy of A. For example, A/I is an A-algebra via its quotient map, which is usually not injective. When k is a field and B is a nonzero k-algebra, the structure map is injective: its kernel is an ideal of the field, and it cannot contain 1.

Definition 6.2. An A-algebra B is *of finite type*, or *finitely generated as an A-algebra*, if there exist $b_1, \ldots, b_n$ such that every element of B is a polynomial in these elements with coefficients in the image of A. Equivalently, the evaluation map

$$A[x_1, \ldots, x_n] \longrightarrow B, \qquad x_i \longmapsto b_i$$

is surjective. It is *finite over A* if it is finitely generated as an A-module.

Module generation allows only linear combinations; algebra generation allows all polynomial expressions. These are different requirements.

Example 6.3 (Finite type need not mean finite)**.** The ring $k[x]$ is generated as a k-algebra by x, but it is not a finite-dimensional k-vector space: the monomials $1, x, x^2, \dots$ are linearly independent. In contrast, $k[x]/(x^2)$ is finite over k, with basis $1, \bar{x}$. More generally, if $g \in A[x]$ is monic of degree $d \geq 1$, division by g gives a unique representative of degree less than d in every residue class. Thus

$$A[x]/(g) \simeq A \oplus A\bar{x} \oplus \cdots \oplus A\bar{x}^{d-1}$$

as an A-module. To see this without dividing by a possibly noninvertible coefficient, subtract $ax^{n-d}g$ from a polynomial with leading term ax^n, for $n \geq d$. Because the leading coefficient of g is 1, this cancels the leading term and lowers the degree. Repeating terminates. Uniqueness follows because a nonzero multiple qg has degree $\deg q + d$: its leading coefficient is the nonzero leading coefficient of q. Thus no such multiple can have degree less than d.

6.2 Restriction of scalars

Fix a ring homomorphism $f : A \to B$. If N is a B-module, we can regard it as an A-module by

$$a \cdot n = f(a)n.$$

This procedure is called *restriction of scalars*. It changes the permitted scalar action without changing the elements or the addition law of N. Every B-linear map is automatically A-linear, because $g(f(a)n) = f(a)g(n)$.

For example, a complex vector space can be regarded as a real vector space using $\mathbb{R} \hookrightarrow \mathbb{C}$. A one-dimensional complex vector space has real dimension two: if e is a complex basis vector, e and ie form a real basis. However, a real-linear map need not be complex-linear. Complex conjugation satisfies $\overline{rz} = r\bar{z}$ for $r \in \mathbb{R}$, but it sends iz to $-i\bar{z}$.

Restriction preserves exact sequences. Indeed, kernels, images, and the underlying maps are unchanged as sets and additive groups, so the equalities defining exactness remain true. Restriction need not preserve finite generation: $\mathbb{Q}$ is generated by 1 as a module over itself, but it is not finitely generated over $\mathbb{Z}$. A finite list of rational numbers has a common denominator d, so its integer span is contained in $d^{-1}\mathbb{Z}$, while $\mathbb{Q}$ is larger.

6.3 Extension of scalars

If M is an A-module, we cannot usually define a B-action directly on the same set of elements. For example, there is no action of $1/2 \in \mathbb{Q}$ on $\mathbb{Z}$ compatible with its usual integer action: such an action would require an integer m with $2m = 1$. The remedy is to construct a new module.

Definition 6.4. The *extension of scalars* of M from A to B is

$$B \otimes_A M,$$

with B-module structure

$$b' \cdot (b \otimes m) = (b'b) \otimes m.$$

The canonical A-linear map is $\eta_M : M \to B \otimes_A M$, $m \mapsto 1 \otimes m$.

Multiplication by b' on the first factor is A-linear, so it induces the displayed operation on the tensor product. The B-module axioms follow by checking pure tensors, which generate the module. The map η_M need not be injective. For $A = \mathbb{Z}$, $B = \mathbb{Q}$, and $M = \mathbb{Z}/2\mathbb{Z}$, every tensor vanishes:

$$b \otimes m = (b/2) \otimes 2m = 0.$$

Thus extension to rational scalars removes this torsion module entirely.

Theorem 6.5 (Extension–restriction correspondence)**.** *For an A-module M and a B-module N, there is a natural bijection*

$$\mathrm{Hom}_B(B \otimes_A M, N) \simeq \mathrm{Hom}_A(M, N),$$

where on the right N is regarded as an A-module by restriction of scalars. Explicitly, a B-linear map F corresponds to

$$g(m) = F(1 \otimes m),$$

and an A-linear map g corresponds to

$$F(b \otimes m) = b\, g(m).$$

Proof. Starting with F, we check that g is A-linear. Additivity is immediate, and

$$g(am) = F(1 \otimes am) = F(f(a) \otimes m) = f(a)F(1 \otimes m) = a \cdot g(m).$$

Conversely, start with g. The formula $(b, m) \mapsto b\, g(m)$ is additive in both variables, and its balancing condition is

$$(bf(a))g(m) = b\big(f(a)g(m)\big) = b\, g(am).$$

It is also A-linear in either variable when the other is fixed. Therefore it induces an A-linear map on the tensor product. This map is B-linear because multiplying b by b' multiplies the result by b'.

The first construction followed by the second recovers F, since $F(b \otimes m) = bF(1 \otimes m)$ by B-linearity. The second followed by the first recovers g by setting $b = 1$. Composing the formulas with other module maps proves naturality. □

This theorem says that $B \otimes_A M$ is the universal way to receive an A-linear map from M into a B-module. The explicit bijection explains both the construction of the extension and its uniqueness for this mapping problem.

6.4 Changing scalars in a presentation

Right exactness immediately gives a practical algorithm. If

$$A^r \xrightarrow{D} A^n \longrightarrow M \longrightarrow 0$$

is a presentation, then

$$B^r \xrightarrow{f(D)} B^n \longrightarrow B \otimes_A M \longrightarrow 0$$

is a presentation after extension. Here $f(D)$ means apply f to every matrix entry. Consequently extension of scalars preserves finite generation and finite presentation. Restriction, as we saw, does not generally preserve finite generation.

Example 6.6 (The same relation over different rings)**.** For $M = \mathbb{Z}/6\mathbb{Z}$, the presentation $\mathbb{Z} \xrightarrow{\times 6} \mathbb{Z} \to M \to 0$ gives

$$B \otimes_{\mathbb{Z}} M \cong \operatorname{coker}(\times 6 : B \to B) = B/6B.$$

Thus

B	$\times 6$ on B	$B \otimes_{\mathbb{Z}} M$
$\mathbb{Q}$	isomorphism	0
$\mathbf{F}_2$	0	$\mathbf{F}_2$
$\mathbf{F}_5$	id	0

The relation $6\bar{1} = 0$ changes with the image of its coefficient 6 in B.

If $A \to B \to C$ are ring maps, scalar extension in two stages agrees with scalar extension in one stage:

$$C \otimes_B (B \otimes_A M) \simeq C \otimes_A M, \quad c \otimes (b \otimes m) \longmapsto cb \otimes m.$$

Here b acts on C through $B \to C$. The inverse is $c \otimes m \mapsto c \otimes (1 \otimes m)$. Balancing verifies both formulas, and the identity $c \otimes (b \otimes m) = cb \otimes (1 \otimes m)$ verifies that the composites are identities. This compatibility lets us simplify repeated scalar changes.

6.5 Tensor products of algebras and equations

If B and C are A-algebras, the module $B \otimes_A C$ has a ring structure determined by

$$(b \otimes c)(b' \otimes c') = bb' \otimes cc', \qquad 1 = 1 \otimes 1.$$

To justify the rule, the expression on the right is linear in each of the four variables and respects every balancing relation used to form either tensor factor. It therefore induces a well-defined bilinear multiplication on the tensor product. Associativity and commutativity hold on pure tensors by the corresponding identities in B, C, and distributivity extends them to arbitrary finite sums. The same argument proves the unit identity. This ring may be zero, which is allowed by our conventions.

The maps $B \to B \otimes_A C$, $b \mapsto b \otimes 1$, and $C \to B \otimes_A C$, $c \mapsto 1 \otimes c$, agree on the images of A. If D is an A-algebra and $g : B \to D$, $h : C \to D$ are A-algebra homomorphisms, then

$$b \otimes c \longmapsto g(b)h(c)$$

defines the unique algebra homomorphism compatible with both. Bilinearity gives the underlying linear map; multiplicativity follows by expanding products of pure tensors. Uniqueness follows because $b \otimes c = (b \otimes 1)(1 \otimes c)$. This is the algebra version of the universal property.

Proposition 6.7 (Changing coefficients of polynomial equations)**.** *Let B be an A-algebra, and let $I = (f_1, \dots, f_r) \subset A[x_1, \dots, x_n]$. Write $f_{i,B}$ for the polynomial obtained by applying $A \to B$ to its coefficients. Then*

$$B \otimes_A \big(A[x_1, \dots, x_n]/I\big) \simeq B[x_1, \dots, x_n]/(f_{1,B}, \dots, f_{r,B})$$

as B-algebras.

Proof. First, $B \otimes_A A[x_1, \dots, x_n] \simeq B[x_1, \dots, x_n]$: the monomials form an A-basis on the left before tensoring, and their tensors form the corresponding B-basis afterward. The map $b \otimes f \mapsto b f_B$ respects products, giving the algebra isomorphism.

Next tensor the exact sequence $I \to A[x_1, \dots, x_n] \to A[x_1, \dots, x_n]/I \to 0$. Right exactness says the desired algebra is the quotient of $B[x_1, \dots, x_n]$ by the image of $B \otimes_A I$. That image is exactly the ideal generated by the $f_{i,B}$: each polynomial combination of the f_i maps to such a combination after changing coefficients, and multiplying $f_{i,B}$ by any monomial and coefficient from B is the image of a corresponding tensor. This identifies the quotient. □

6.6 Fiber algebras

Let k be algebraically closed of characteristic different from 2. Consider $A = k[t]$ and the A-algebra $B = k[x]$ with $t \mapsto x^2$. A choice of parameter value $a \in k$ is represented algebraically by the evaluation homomorphism $k[t] \to k$, $t \mapsto a$. Denote this copy of k, with its specified A-action, by k_a. The corresponding *fiber algebra* is

$$B \otimes_A k_a \simeq k[x]/(x^2 - a).$$

To verify the formula, regard B as $k[t,x]/(x^2 - t)$ and use the preceding proposition. The ordinary points of this fiber are the solutions to $x^2 = a$.

If $a \neq 0$, choose $r \in k$ with $r^2 = a$. The two roots $r, -r$ are distinct, so the ideals $(x - r)$ and $(x + r)$ are comaximal: their difference contains the nonzero constant $2r$. The Chinese remainder theorem gives

$$k[x]/(x^2 - a) \simeq k \times k.$$

At $a = 0$, the fiber algebra is $k[x]/(x^2)$, containing a nonzero nilpotent. It has only one ordinary point, yet it still has k-dimension two. Thus the algebra retains information when two points merge.

The total algebra B is free of rank two over A: every polynomial in x has a unique expression

$$p(x^2) + xq(x^2), \qquad p, q \in k[t].$$

Separate the even and odd powers to see existence and uniqueness. Therefore all fiber algebras here have dimension two, although their numbers of distinct ordinary points differ. This is an elementary example of why the algebra of a family is more informative than a list of its solutions.

A *Laurent polynomial* in x is a finite sum $\sum_{i=m}^{n} a_i x^i$ in which negative exponents are allowed. The ring $k[x, x^{-1}]$ uses the usual sum and product of these expressions and satisfies $xx^{-1} = 1$.

Finally, note the direction of the maps. The rule on ordinary points $x \mapsto x^2$ goes from the x-line to the parameter t-line, whereas substitution sends functions in the opposite direction, $k[t] \to k[x]$, $t \mapsto x^2$. A function on the target can be evaluated after the point map to obtain a function on the source. This reversal is the basic reason ring homomorphisms will describe geometric maps in the opposite direction.

Three fibers of one equation

Let $A = k[t]$ and $B = A[x]/(x^2 - t)$, where $\operatorname{char} k \neq 2$. As an A-module, B is free with basis $1, x$: division by the monic polynomial $x^2 - t$ gives a unique remainder of degree less than two in x. At the closed point $(t - a)$ the fiber ring is

$$B \otimes_A k \cong k[x]/(x^2 - a).$$

If $a = b^2 \neq 0$ in k, the two linear factors are comaximal because $2b$ is a unit. The Chinese remainder theorem gives $k[x]/(x^2 - a) \cong k \times k$. If $a = 0$, the fiber is $k[x]/(x^2)$, whose underlying spectrum has one point but whose vector-space dimension is two. If a is not a square in k, the polynomial is irreducible and the fiber is a quadratic field extension of k. All three fibers have dimension two as k-vector spaces. Their point sets and nilpotents are different.

The generic point of $\operatorname{Spec} k[t]$ is (0), with residue field $k(t)$. Its fiber is

$$k(t)[x]/(x^2 - t).$$

Here t is not a square in $k(t)$: the exponent of t in a square rational function is even, whereas its exponent in t is one. Thus the generic fiber is a field. This is a first example of how a family can be studied over a function field.

Base change and matrices

If M is presented by $A^r \xrightarrow{D} A^s \to M \to 0$ and $A \to C$ is a ring map, then

$$C^r \xrightarrow{D_C} C^s \longrightarrow C \otimes_A M \longrightarrow 0$$

is a presentation, where D_C is obtained by applying $A \to C$ to each matrix entry. For example, take $M = \operatorname{coker}(A \xrightarrow{t} A)$. The fiber at $t = a$ is the cokernel of multiplication by a on k. It is zero for $a \neq 0$ and k for $a = 0$. Although the original presentation has the same sizes at every point, its matrix rank changes. Right exactness guarantees the cokernel calculation; it does not guarantee preservation of the original kernel.

6.7 Finiteness under changing scalars

The two next assertions use different generating sets. Restricting scalars makes fewer coefficients available; extending scalars makes more coefficients available. Keeping that distinction in view prevents a common confusion between finite modules and finitely generated algebras.

Proposition 6.8 (Restriction and finite generation; AM 2.16)**.** *Let $A \to B$ be a ring homomorphism. If B is finitely generated as an A-module and a B-module N is finitely generated over B, then N is finitely generated as an A-module.*

Proof. Choose $B = \sum_{i=1}^{r} Ab_i$ and $N = \sum_{j=1}^{s} Bn_j$. Writing A-coefficients through the structure map,

$$n = \sum_j c_j n_j, \qquad c_j = \sum_i a_{ij} b_i \quad \Longrightarrow \quad n = \sum_{i,j} a_{ij}(b_i n_j).$$

Hence $N = \sum_{i,j} A(b_i n_j)$ has at most rs generators over A. □

Proposition 6.9 (Extension and finite generation; AM 2.17)**.** *If M is finitely generated over A, then $B \otimes_A M$ is finitely generated over B.*

Proof. Write $M = \sum_{i=1}^{r} Am_i$ and let $f : A \to B$ be the structure map. Balancing gives

$$m = \sum_i a_i m_i \quad \Longrightarrow \quad b \otimes m = \sum_i b f(a_i)(1 \otimes m_i).$$

Every tensor is a finite sum of pure tensors. Thus, for arbitrary B,

$$B \otimes_A M = \sum_{i=1}^{r} B(1 \otimes m_i).$$

□

Example 6.10. The algebra $B = \mathbb{Z}[i]$ is generated by $1, i$ as a $\mathbb{Z}$-module. Thus a B-module generated by r elements is generated by at most $2r$ elements over $\mathbb{Z}$. In contrast, $\mathbb{Z}[x]$ is generated by one element as a $\mathbb{Z}$-*algebra* but requires infinitely many elements as a $\mathbb{Z}$-module. One cannot replace the first hypothesis of Proposition 6.8 by "finitely generated algebra": take $N = B = \mathbb{Z}[x]$.

6.8 Tensor products over two rings

Sometimes two commuting scalar actions come from different rings. If N is both an A-module and a B-module, the actions are *compatible* when $a(bn) = b(an)$ for all a, b, n. We may write the B-action on the right, $nb = bn$, to distinguish it from the A-action. A module with these two compatible actions is called an (A,B)-*bimodule*. The following result specifies the ring used for each tensor product.

Proposition 6.11 (Mixed associativity; AM in-text Exercise 2.15)**.** *Let M be an A-module, let P be a B-module, and let N carry compatible A- and B-module structures, so that $a(nb) = (an)b$. The formulas $b(m \otimes n) = m \otimes nb$ and $a(n \otimes p) = an \otimes p$ make $M \otimes_A N$ a B-module and $N \otimes_B P$ an A-module, respectively. There is a canonical isomorphism*

$$(M \otimes_A N) \otimes_B P \cong M \otimes_A (N \otimes_B P), \qquad (m \otimes n) \otimes p \longmapsto m \otimes (n \otimes p).$$

Proof. Define $b(m \otimes n) = m \otimes nb$. The map $(m,n) \mapsto m \otimes nb$ respects the A-balancing relation because the two actions on N commute; hence it induces an endomorphism of $M \otimes_A N$. The identities for a B-module follow on the generating pure tensors. Likewise define $a(n \otimes p) = an \otimes p$; compatibility makes it well defined over B.

To construct the displayed map, start with the rule on triples $(m,n,p) \mapsto m \otimes (n \otimes p)$. It is additive in all variables and satisfies the two balancing rules

$$(am, n, p) \sim (m, an, p), \qquad (m, nb, p) \sim (m, n, bp).$$

First enforce the A-balancing relation to obtain a map from $(M \otimes_A N) \times P$; then enforce the B-balancing relation to obtain the desired map. The reverse rule $m \otimes (n \otimes p) \mapsto (m \otimes n) \otimes p$ respects exactly the same relations in the reverse order. Both composites are the identity on all generating triple tensors, so the maps are inverse. □

This proof also gives the extension-of-scalars identity

$$C \otimes_B (B \otimes_A M) \cong C \otimes_A M$$

for homomorphisms $A \to B \to C$. The actual maps are $c \otimes (b \otimes m) \mapsto cb \otimes m$ and $c \otimes m \mapsto c \otimes (1 \otimes m)$.

Proposition 6.12 (Flatness after base change; AM in-text Exercise 2.20)**.** *If M is flat over A, then $B \otimes_A M$ is flat over B.*

Proof. Take an injection $L \to N$ of B-modules. It is still injective after restriction to A, because the underlying function has not changed. Flatness of M gives an injection $L \otimes_A M \to N \otimes_A M$. For any B-module L there is an isomorphism

$$L \otimes_B (B \otimes_A M) \longrightarrow L \otimes_A M, \quad l \otimes (b \otimes m) \longmapsto bl \otimes m,$$

with inverse $l \otimes m \mapsto l \otimes (1 \otimes m)$. The balancing checks are the ones just proved. These identifications commute with the map $L \to N$, so the injection is exactly the map obtained by tensoring over B with $B \otimes_A M$. This proves flatness by the injection criterion. □

An A-algebra B is called *flat over* A if its underlying A-module is flat. Thus the term refers to preservation of injections under tensoring, and does not impose an additional condition on multiplication.

6.9 Exercises

Exercise 6.1 (Compare two notions of generation)**.** Prove that every finite A-algebra is of finite type. Show that $\mathbb{Z}[1/2]$ is of finite type over $\mathbb{Z}$ but is not finite over $\mathbb{Z}$.

Exercise 6.2 (Extend a map by its values)**.** Let $M = \mathbb{Z}^2$ and let N be a complex vector space. Describe every complex-linear map $\mathbb{C} \otimes_{\mathbb{Z}} M \to N$. What is $\mathbb{C} \otimes_{\mathbb{Z}} M$?

Exercise 6.3 (Extension and contraction of ideals)**.** For $f : A \to B$ and an ideal $I \subseteq A$, let IB be the ideal generated by $f(I)$. Prove $B \otimes_A A/I \simeq B/IB$. Is $f^{-1}(IB) = I$ always true?

Exercise 6.4 (The family $xy = t$)**.** Let $A = k[t]$ and $B = k[x,y]$, with $t \mapsto xy$. Compute the fiber algebra at $a \in k$. Show that for $a \neq 0$ it is isomorphic to $k[x,x^{-1}]$, the ring of finite Laurent polynomials in x, while the fiber at 0 has zero divisors and is reduced.

Exercise 6.5 (Polynomials with module coefficients; AM Chapter 2, Exercise 6)**.** Let M be an A-module, and let $M[x]$ consist of finite formal sums $\sum_{i=0}^{r} m_i x^i$, with $m_i \in M$. Make it an $A[x]$-module and prove $M[x] \cong A[x] \otimes_A M$ as $A[x]$-modules.

Exercise 6.6 (Two permanence properties for flat modules; AM Chapter 2, Exercise 8)**.**

1. If M, N are flat over A, prove that $M \otimes_A N$ is flat over A.
2. If B is a flat A-algebra and N is a flat B-module, prove that N is flat as an A-module.

Lecture 7
Localization: choosing which elements may be denominators

Localization allows specified elements of a ring to become denominators. Its construction must account for zero divisors, and this is the source of the extra factor in the equality test for fractions. The examples $\mathbb{Z}$, $k[x]$, and $k[x,y]/(xy)$ illustrate the different possibilities. Reading: [1, Chapter 3].

7.1 Multiplicative subsets and rings of fractions

Passing from $\mathbb{Z}$ to $\mathbb{Q}$ inverts every nonzero integer. Two choices of fewer denominators are

$$\mathbb{Z}[1/2] = \left\{ \frac{a}{2^r} : a \in \mathbb{Z},\ r \geq 0 \right\}, \qquad \mathbb{Z}_{(3)} = \left\{ \frac{a}{b} : a,b \in \mathbb{Z},\ 3 \nmid b \right\}.$$

For example,

	$\mathbb{Z}[1/2]$	$\mathbb{Z}_{(3)}$
$1/2$	$\in$	$\in$
$1/5$	$\notin$	$\in$
$1/3$	$\notin$	$\notin$

The first construction inverts 2; the second inverts every integer prime to 3.

Definition 7.1. A subset $S \subseteq A$ is *multiplicatively closed* if $1 \in S$ and $s,t \in S$ imply $st \in S$. The *localization* $S^{-1}A$ is the ring obtained by allowing the elements of S as denominators.

The last sentence is motivation, not yet a construction. We construct it using pairs (a,s) with $a \in A$ and $s \in S$. Declare

$$(a,s) \sim (b,t) \quad \Longleftrightarrow \quad u(at-bs) = 0 \text{ for some } u \in S.$$

Write a/s for the equivalence class of (a,s). The extra factor u is essential in a ring with zero divisors: making u invertible forces every element annihilated by u to become zero.

Let us check the least obvious part, transitivity. Suppose $u(at-bs) = 0$ and $v(br-ct) = 0$. Multiplying the first equation by vr, the second by us, and adding gives

$$uvt(ar-cs)=0.$$

Since $uvt \in S$, this proves $(a,s)\sim(c,r)$. Reflexivity uses $u=1$, and symmetry follows by changing a sign.

Define the ring operations by

$$\frac{a}{s}+\frac{b}{t}=\frac{at+bs}{st}, \qquad \frac{a}{s}\frac{b}{t}=\frac{ab}{st}.$$

To see why these do not depend on representatives, suppose first that $u(as'-a's)=0$ and $v(bt'-b't)=0$. For products, use

$$abs't'-a'b'st=bt'(as'-a's)+a's(bt'-b't).$$

The right side is killed by uv. For sums with denominators st and $s't'$, the cross difference is

$$tt'(as'-a's)+ss'(bt'-b't),$$

again killed by uv. The ordinary distributive and associative calculations then hold for these classes. The elements $0/1$ and $1/1$ are the additive and multiplicative identities.

Proposition 7.2 (The zero test)**.** *For $a \in A$ and $s \in S$,*

$$\frac{a}{s}=0 \text{ in } S^{-1}A \quad\Longleftrightarrow\quad ua=0 \text{ for some } u\in S.$$

Consequently, the map $A \to S^{-1}A$, $a \mapsto a/1$, need not be injective.

Proof. Apply the equivalence relation to (a,s) and $(0,1)$. Their cross difference is exactly a. Notice that the particular denominator s disappears from the criterion. □

For $S=\{1,f,f^2,\ldots\}$ write A_f. For a prime ideal $\mathfrak{p}$, the complement $A\setminus\mathfrak{p}$ is multiplicatively closed: if neither a nor b lies in $\mathfrak{p}$, primality says $ab\notin\mathfrak{p}$. We write

$$A_{\mathfrak{p}}=(A\setminus\mathfrak{p})^{-1}A.$$

These notations must not be confused. A_f inverts one element and all its powers; $A_{\mathfrak{p}}$ inverts every element outside a prime.

When A is an integral domain and $S=A\setminus\{0\}$, the localization is a field, denoted $\mathrm{Frac}(A)$ and called the *fraction field* of A. Indeed, the zero test shows that $a/s\neq 0$ exactly when $a\neq 0$, and then s/a is its inverse. The map $A\to\mathrm{Frac}(A)$ is injective, because no nonzero element of a domain is killed by a nonzero scalar. For example, $\mathrm{Frac}(\mathbb{Z})=\mathbb{Q}$.

Example 7.3. Let $A=k[x,y]/(xy)$ and invert x. Since $xy=0$ and x becomes a unit, y becomes zero. In fact

$$A_x\cong k[x,x^{-1}].$$

One map sends x to x and y to 0; its inverse sends the Laurent variable x to the class $x/1$. These maps are inverse on the displayed generators. Localization can therefore remove an entire part of a ring, not merely add symbols for fractions.

Why equality of fractions needs an extra factor

Take $A = \mathbb{Z}/6\mathbb{Z}$ and invert 2. Since $2 \cdot 3 = 0$, inverting 2 forces $3 = 0$. Thus $0/1 = 3/1$ in $A[1/2]$, even though $0 \neq 3$ in A. The equality criterion is

$$a/s = b/t \quad \Longleftrightarrow \quad u(ta - sb) = 0 \text{ for some } u \in S,$$

not simply $ta = sb$. In this example $u = 2$ witnesses the equality. Indeed $A[1/2] \cong \mathbb{Z}/3\mathbb{Z}$: the quotient map sends 2 to a unit, and in the localization the relation $3 = 0$ gives a map in the other direction. The maps are inverse on all fractions.

Transitivity of the fraction relation is also worth checking. If

$$u(ta - sb) = 0, \qquad v(rb - tc) = 0,$$

then multiplying the first equality by vr and the second by us, and adding, gives

$$uvt(ra - sc) = 0.$$

Since $uvt \in S$, this proves $a/s = c/r$. The factor t cannot in general be cancelled inside A; it is absorbed into the allowed witness.

7.2 The universal property

Theorem 7.4 (Universal property; AM 3.1)**.** *Suppose $f : A \to B$ is a ring homomorphism such that every $f(s)$, $s \in S$, is a unit. There is a unique homomorphism*

$$\widetilde{f} : S^{-1}A \longrightarrow B, \qquad \widetilde{f}(a/s) = f(a)f(s)^{-1},$$

whose composite with $A \to S^{-1}A$ is f.

Proof. Uniqueness follows from $a/s = (a/1)(s/1)^{-1}$. For well-definedness,

$$\begin{aligned} \frac{a}{s} = \frac{b}{t} &\Longrightarrow u(at - bs) = 0 \quad \text{for some } u \in S \\ &\Longrightarrow f(u)\big(f(a)f(t) - f(b)f(s)\big) = 0 \\ &\Longrightarrow f(a)f(s)^{-1} = f(b)f(t)^{-1}, \end{aligned}$$

where $f(u), f(s), f(t)$ are units. The arithmetic checks are

$$\begin{aligned} \widetilde{f}\left(\frac{at + bs}{st}\right) &= f(a)f(s)^{-1} + f(b)f(t)^{-1}, \\ \widetilde{f}\left(\frac{ab}{st}\right) &= f(a)f(s)^{-1}f(b)f(t)^{-1}, \\ \widetilde{f}(1/1) &= 1. \end{aligned}$$

Thus $\widetilde{f}$ is a unital ring homomorphism extending f. $\square$

The practical use is to construct maps *out of* a localization without checking every fraction separately. Specify the map on A, verify that the permitted denominators go to units, and the extension follows.

7.3 Modules and exactness

For an A-module M, define $S^{-1}M$ by fractions m/s, using

$$\frac{m}{s} = \frac{n}{t} \quad \Longleftrightarrow \quad u(tm - sn) = 0 \text{ for some } u \in S.$$

Addition and scalar multiplication are

$$\frac{m}{s} + \frac{n}{t} = \frac{tm + sn}{st}, \qquad \frac{a}{s}\frac{m}{t} = \frac{am}{st}.$$

The same common-denominator calculations used for $S^{-1}A$ show that these formulas are independent of representatives and satisfy the module axioms. The canonical map is $M \to S^{-1}M$, $m \mapsto m/1$. For $S = \{1, f, f^2, \ldots\}$ we write M_f; for $S = A \setminus \mathfrak{p}$ we write $M_{\mathfrak{p}}$. The zero test becomes $m/s = 0$ if and only if some element of S kills m. A homomorphism $h : M \to N$ localizes to $S^{-1}h(m/s) = h(m)/s$.

Theorem 7.5 (Exactness of localization; AM 3.3)**.** *If*

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

is exact at M, then its localization is exact at $S^{-1}M$.

Proof. Since $(S^{-1}g)(S^{-1}f) = S^{-1}(gf) = 0$, we have $\operatorname{im}(S^{-1}f) \subseteq \ker(S^{-1}g)$. Conversely,

$$\begin{aligned} \frac{m}{s} \in \ker(S^{-1}g) &\Longrightarrow \frac{g(m)}{s} = 0 \\ &\Longrightarrow ug(m) = g(um) = 0 \quad \text{for some } u \in S \\ &\Longrightarrow um = f(m') \quad \text{for some } m' \in M'. \end{aligned}$$

Therefore

$$\frac{m}{s} = \frac{um}{us} = S^{-1}f\left(\frac{m'}{us}\right) \in \operatorname{im}(S^{-1}f).$$

The extra denominator u allows us to replace m by $um \in \ker g$. $\square$

In particular, injections and surjections remain so after localization, and

$$S^{-1}(M/N) \cong S^{-1}M / S^{-1}N.$$

Localization is also an example of tensor product:

$$S^{-1}A \otimes_A M \xrightarrow{\sim} S^{-1}M, \qquad (a/s) \otimes m \longmapsto am/s.$$

Its inverse is $m/s \mapsto (1/s) \otimes m$. If $u(tm - sn) = 0$, then the difference of these two proposed tensor expressions equals $(1/(ust)) \otimes u(tm - sn) = 0$; hence the inverse is well-defined. This proves that $S^{-1}A$ is a flat A-module: tensoring with it preserves exact sequences.

7.4 Prime ideals and local tests

If I is an ideal of A, its extension to $S^{-1}A$ consists exactly of the fractions with numerator in I. We denote it by $S^{-1}I$ or $IS^{-1}A$. One inclusion follows from $i/s = (i/1)(1/s)$; the other follows by writing a finite sum of ideal generators over a common denominator. Moreover,

$$\frac{a}{s} \in S^{-1}I \quad \Longleftrightarrow \quad ua \in I \text{ for some } u \in S.$$

For necessity, write $a/s = i/t$ and choose $v \in S$ with $v(ta - si) = 0$, giving $vta \in I$. For sufficiency, write $a/s = (ua)/(us)$.

In an expression $S^{-1}(A/I)$, the denominators mean the images of the elements of S in A/I. The quotient isomorphism above sends $(a+I)/s$ to $a/s + S^{-1}I$. Its formula respects multiplication, so it is also an isomorphism of rings.

Theorem 7.6 (Prime correspondence)**.** *Extension and contraction give inverse, inclusion-preserving bijections*

$$\{\mathfrak{p} \in \operatorname{Spec} A : \mathfrak{p} \cap S = \varnothing\} \longleftrightarrow \operatorname{Spec}(S^{-1}A), \qquad \mathfrak{p} \longmapsto S^{-1}\mathfrak{p}.$$

Proof. For $\mathfrak{p} \cap S = \varnothing$, the quotient

$$S^{-1}A/S^{-1}\mathfrak{p} \cong S^{-1}(A/\mathfrak{p})$$

is a nonzero domain; hence $S^{-1}\mathfrak{p}$ is prime. Its contraction is $\mathfrak{p}$, because

$$\begin{aligned} a/1 \in S^{-1}\mathfrak{p} &\Longleftrightarrow ua \in \mathfrak{p} \text{ for some } u \in S \\ &\Longleftrightarrow a \in \mathfrak{p} \qquad (u \notin \mathfrak{p}). \end{aligned}$$

Conversely, let $\mathfrak{p} = \{a \in A : a/1 \in Q\}$ for a prime Q of $S^{-1}A$. Inverse images preserve primality, and

$$s \in S \implies s/1 \text{ is a unit} \implies s/1 \notin Q.$$

Thus $\mathfrak{p} \cap S = \varnothing$. Since $s/1$ is a unit,

$$a/s \in Q \iff a/1 \in Q \iff a \in \mathfrak{p},$$

which gives $Q = S^{-1}\mathfrak{p}$. □

Definition 7.7 (Local rings and residue fields)**.** A nonzero ring A is *local* if it has a unique maximal ideal $\mathfrak{m}$. The quotient $\kappa = A/\mathfrak{m}$ is its *residue field*. We often write $(A, \mathfrak{m}, \kappa)$ to record all three objects.

For $A_{\mathfrak{p}}$, the surviving primes are exactly the primes contained in $\mathfrak{p}$. In particular, $\mathfrak{p}A_{\mathfrak{p}}$ is the unique maximal ideal. Thus $A_{\mathfrak{p}}$ is a local ring. Its residue field is denoted $\kappa(\mathfrak{p}) = A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$. For $A = \mathbb{Z}$ and $\mathfrak{p} = (p)$, it is $\mathbf{F}_p$; for $\mathfrak{p} = (0)$ in a domain, the local ring and its residue field both equal $\mathrm{Frac}(A)$.

Proposition 7.8 (Detecting zero locally)**.** *An A-module M is zero if and only if $M_{\mathfrak{m}} = 0$ for every maximal ideal $\mathfrak{m}$. No finite-generation hypothesis is needed.*

Proof. If $0 \neq m \in M$, then $I = \mathrm{Ann}_A(m) \neq A$. Choose a maximal ideal $\mathfrak{m} \supseteq I$. The zero test gives

$$m/1 = 0 \text{ in } M_{\mathfrak{m}} \implies sm = 0 \text{ for some } s \notin \mathfrak{m} \implies s \in I \subseteq \mathfrak{m},$$

a contradiction. Hence $M \neq 0$ implies $M_{\mathfrak{m}} \neq 0$ for some maximal ideal. The converse implication is immediate. □

For a map $h : M \to N$, exactness of localization gives

$$(\ker h)_{\mathfrak{m}} = \ker h_{\mathfrak{m}}, \qquad (\mathrm{coker}\, h)_{\mathfrak{m}} \cong \mathrm{coker}\, h_{\mathfrak{m}}.$$

The zero criterion therefore yields

$$h \text{ is an isomorphism} \iff h_{\mathfrak{m}} \text{ is an isomorphism for every maximal } \mathfrak{m}.$$

For $L \xrightarrow{u} M \xrightarrow{v} N$, suppose every localization is exact. Then

$$(\mathrm{im}(vu))_{\mathfrak{m}} = \mathrm{im}(v_{\mathfrak{m}} u_{\mathfrak{m}}) = 0 \implies vu = 0.$$

Now $H = \ker v / \mathrm{im}\, u$ is defined, and

$$H_{\mathfrak{m}} \cong \ker v_{\mathfrak{m}} / \mathrm{im}\, u_{\mathfrak{m}} = 0 \implies H = 0.$$

Hence $\mathrm{im}\, u = \ker v$. Applying this argument at each term proves that exactness can be tested at all maximal ideals.

A local calculation can discard an entire component

Let $A = k[x,y]/(xy)$. At the prime $\mathfrak{p} = (x)$, the element y is inverted, so the equation $xy = 0$ implies $x = 0$. More precisely,

$$A_{(x)} \cong k(y).$$

To see this, reduce modulo x. Every denominator outside (x) becomes a nonzero polynomial in y, and every such polynomial is allowed. This produces the fraction field of $k[y]$. The map has no remaining kernel because x has already become zero.

At $\mathfrak{m} = (x, y)$ neither x nor y is inverted. Both survive as nonzero zero divisors. For instance, if $x/1 = 0$, some $s \notin \mathfrak{m}$ would satisfy $sx = 0$. But $\mathrm{Ann}_A(x) = (y) \subseteq \mathfrak{m}$, a contradiction. Thus $A_{\mathfrak{m}}$ remembers both axes at their intersection, whereas $A_{(x)}$ remembers the generic behavior of the component $x = 0$.

This also distinguishes two uses of the word local. Localization at a prime is an algebraic construction with a specified multiplicative set. A property being local means it can be tested on a family of these constructions. The theorem that vanishing is local uses an annihilator and a maximal ideal; it does not assume finite generation of the module.

7.5 Recognizing and combining localizations

Corollary 7.9 (A recognition criterion; AM 3.2)**.** *Let $g : A \to B$ satisfy the following conditions for a multiplicatively closed set S:*

1. *$g(s)$ is a unit for every $s \in S$;*
2. *if $g(a) = 0$, then $sa = 0$ for some $s \in S$;*
3. *every element of B equals $g(a)g(s)^{-1}$ for some $a \in A, s \in S$.*

There is a unique isomorphism $S^{-1}A \to B$ commuting with the maps from A.

Proof. The universal property gives the unique map $h(a/s) = g(a)g(s)^{-1}$. Condition (3) says it is onto. If $h(a/s) = 0$, multiplying by the unit $g(s)$ gives $g(a) = 0$. By (2), some $t \in S$ kills a, and the zero test gives $a/s = 0$. Thus h is injective as well. □

7.6 Localization of module operations

Corollary 7.10 (Sums, intersections, and quotients; AM 3.4)**.** *For submodules N, P of M,*

$$S^{-1}(N+P) = S^{-1}N + S^{-1}P,$$
$$S^{-1}(N \cap P) = S^{-1}N \cap S^{-1}P,$$
$$S^{-1}(M/N) \cong (S^{-1}M)/(S^{-1}N).$$

The first two equalities use the injective maps of localized submodules into $S^{-1}M$. The corresponding assertions hold for any finite sum or intersection.

Proof. Exactness of localization permits the stated identifications of submodules. For sums, $(n+p)/s = n/s + p/s$, while $n/s + p/t = (tn+sp)/(st)$ lies in the localization of the sum. This proves both inclusions.

For intersections, one inclusion is immediate. For the other, represent an element in the intersection as $n/s = p/t$, with $n \in N, p \in P$. Some $u \in S$ satisfies $u(tn - sp) = 0$. Hence $w = utn = usp$ lies in $N \cap P$, and $n/s = w/(uts)$. Thus the element comes from the localized intersection.

For the quotient, localize $0 \to N \to M \to M/N \to 0$. Exactness and the first isomorphism theorem give the isomorphism, explicitly $(m+N)/s \mapsto m/s + S^{-1}N$. Repeating the two-factor assertions gives any finite number of factors. □

Caution 7.11. Infinite intersections need not commute with localization. In $\mathbb{Z}$, let $I_n = 2^n\mathbb{Z}$ and invert 2. Their intersection in $\mathbb{Z}$ is zero, but every localized I_n equals $\mathbb{Z}[1/2]$, so the intersection of their localizations is the whole ring.

Proposition 7.12 (Localization is extension of scalars; AM 3.5)**.** *For every A-module M, there is a unique $S^{-1}A$-linear isomorphism*

$$S^{-1}A \otimes_A M \longrightarrow S^{-1}M, \qquad (a/s) \otimes m \longmapsto am/s.$$

Proof. The formula is induced by an A-bilinear map. It is $S^{-1}A$-linear since multiplying a/s before forming the image agrees with multiplying the resulting fraction. An inverse is $m/s \mapsto (1/s) \otimes m$. To verify it, if $m/s = n/t$ and $u(tm - sn) = 0$, the difference of the proposed images is

$$(1/s) \otimes m - (1/t) \otimes n = (1/(ust)) \otimes u(tm - sn) = 0.$$

Thus it is well defined. The composites fix fractions and pure tensors; for the latter use $(1/s) \otimes am = (a/s) \otimes m$. Uniqueness follows because pure tensors generate. □

Corollary 7.13 (Flatness of a ring of fractions; AM 3.6)**.** *The A-module $S^{-1}A$ is flat.*

Proof. Under Proposition 7.12, tensoring with $S^{-1}A$ is localization. Localization preserves exact sequences by Theorem 7.5; in particular it preserves injections, which is the flatness criterion. □

Proposition 7.14 (Tensor products commute with localization; AM 3.7)**.** *There is a unique $S^{-1}A$-linear isomorphism*

$$S^{-1}M \otimes_{S^{-1}A} S^{-1}N \cong S^{-1}(M \otimes_A N), \quad (m/s) \otimes (n/t) \longmapsto (m \otimes n)/(st).$$

In particular, $(M \otimes_A N)_{\mathfrak{p}} \cong M_{\mathfrak{p}} \otimes_{A_{\mathfrak{p}}} N_{\mathfrak{p}}$ for every prime $\mathfrak{p}$.

Proof. Write $B = S^{-1}A$. Substitute $S^{-1}M = B \otimes_A M$ and $S^{-1}N = B \otimes_A N$. The mixed associativity identity gives

$$(B \otimes_A M) \otimes_B (B \otimes_A N) \cong B \otimes_A (M \otimes_A N).$$

The map sends $(b \otimes m) \otimes (b' \otimes n)$ to $bb' \otimes (m \otimes n)$, with inverse $b \otimes (m \otimes n) \mapsto (b \otimes m) \otimes (1 \otimes n)$. Balancing over B moves b' from one outer factor to the other, proving that the two formulas are inverse. Applying Proposition 7.12 again gives the stated fraction formula. Its values on generating tensors determine it uniquely. □

7.7 Local criteria

Proposition 7.15 (Vanishing is local; AM 3.8)**.** *For any A-module M, the following are equivalent: $M = 0$; $M_{\mathfrak{p}} = 0$ for every prime $\mathfrak{p}$; $M_{\mathfrak{m}} = 0$ for every maximal ideal $\mathfrak{m}$.*

Proof. The first condition implies the second, and the second implies the third because maximal ideals are prime. For the converse, if $0 \neq x \in M$, its proper annihilator lies in a maximal ideal $\mathfrak{m}$. The equation $x/1 = 0$ in $M_{\mathfrak{m}}$ would require an element outside $\mathfrak{m}$ annihilating x, a contradiction. Thus $M_{\mathfrak{m}} \neq 0$ for this $\mathfrak{m}$. □

Proposition 7.16 (Injectivity and surjectivity are local; AM 3.9)**.** *A module map $f : M \to N$ is injective if and only if $f_{\mathfrak{p}}$ is injective for every prime $\mathfrak{p}$, if and only if $f_{\mathfrak{m}}$ is injective for every maximal $\mathfrak{m}$. The three analogous conditions with "surjective" are also equivalent.*

Proof. Exactness of localization identifies $(\ker f)_{\mathfrak{p}}$ with $\ker f_{\mathfrak{p}}$, and $(\operatorname{coker} f)_{\mathfrak{p}}$ with $\operatorname{coker} f_{\mathfrak{p}}$. Injectivity means vanishing of the kernel and surjectivity means vanishing of the cokernel. Apply Proposition 7.15 separately to these two modules. This proves both sets of equivalences without a finite-generation assumption. □

Proposition 7.17 (Flatness is local; AM 3.10)**.** *For an A-module M, the following are equivalent: M is flat over A; $M_{\mathfrak{p}}$ is flat over $A_{\mathfrak{p}}$ for every prime $\mathfrak{p}$; $M_{\mathfrak{m}}$ is flat over $A_{\mathfrak{m}}$ for every maximal $\mathfrak{m}$.*

Proof. If M is flat, extension of scalars preserves flatness by Proposition 6.12; applying it to $A \to A_{\mathfrak{p}}$ proves the second condition. That condition implies the third.

Conversely suppose all $M_{\mathfrak{m}}$ are flat, and take an injection $L \to N$. Localizing gives an injection $L_{\mathfrak{m}} \to N_{\mathfrak{m}}$. Tensoring it over $A_{\mathfrak{m}}$ with the flat module $M_{\mathfrak{m}}$ remains injective. Proposition 7.14 identifies the result with the localization of $L \otimes_A M \to N \otimes_A M$. By Proposition 7.16, this original map is injective because it is injective at every maximal ideal. Thus M is flat. □

7.8 Extension and contraction under localization

For the canonical map $A \to S^{-1}A$, write $I^e = S^{-1}I$ for extension of an ideal, and $J^c = \{a : a/1 \in J\}$ for contraction. An ideal of A is called *contracted* if it equals J^c for some ideal J of the localization.

Proposition 7.18 (Extended and contracted ideals, all parts; AM 3.11)**.**

1. *Every ideal of $S^{-1}A$ is extended: $J = (J^c)^e$.*
2. *$I^{ec} = \bigcup_{s\in S}(I : s)$. In particular $I^e = S^{-1}A$ if and only if $I \cap S \neq \varnothing$.*
3. *An ideal I is contracted if and only if multiplication by each $s \in S$ is injective on A/I.*
4. *Extension and contraction give inverse bijections between primes of A disjoint from S and primes of $S^{-1}A$.*
5. *Localization commutes with finite sums, products, intersections, and radicals of ideals. In particular $S^{-1}\sqrt{I} = \sqrt{S^{-1}I}$.*

Proof. For (1), if $a/s \in J$, multiply by $s/1$ to see $a/1 \in J$. Thus $a \in J^c$ and $a/s \in (J^c)^e$. The opposite inclusion follows since J contains all fractions with numerator in J^c.

For (2), $a \in I^{ec}$ means $a/1 = i/s$ for some $i \in I$ and $s \in S$. Some $t \in S$ then satisfies $t(sa - i) = 0$, so $tsa \in I$. Conversely, if $sa \in I$, then $a/1 = (sa)/s \in I^e$. These implications give exactly the union formula. Applying it to $a = 1$ proves the unit-ideal assertion.

For (3), an ideal is contracted precisely when $I = I^{ec}$. Indeed (1) implies $(J^c)^{ec} = J^c$, and an ideal fixed by extension followed by contraction is itself contracted. By (2), $I = I^{ec}$ means that $sa \in I$ with $s \in S$ always implies $a \in I$. This is exactly injectivity of multiplication by s on A/I. This formulation includes the zero quotient without any ambiguity about zero divisors.

For (4), a prime $\mathfrak{p}$ disjoint from S has $S^{-1}(A/\mathfrak{p})$ a nonzero domain, so $S^{-1}\mathfrak{p}$ is prime; by primality $sa \in \mathfrak{p}$, $s \notin \mathfrak{p}$ imply $a \in \mathfrak{p}$, so its contraction is $\mathfrak{p}$. Conversely the contraction of a prime is prime and cannot meet S, since all its elements become units. Part (1) recovers the original prime by extension.

For (5), sums and intersections were proved in Corollary 7.10. Products commute because a product $(a/s)(b/t) = ab/(st)$ has numerator in IJ when $a \in I, b \in J$, and conversely $(ab)/s = (a/1)(b/s)$ lies in the product of the extended ideals. Finite sums of such products prove equality for IJ; induction handles more factors.

For radicals, if $a \in \sqrt{I}$, a power of a/s lies in $S^{-1}I$, proving one inclusion. Conversely, if $(a/s)^n \in S^{-1}I$, membership in an extended ideal yields $ta^n \in I$ for some $t \in S$. Then $(ta)^n = t^{n-1}(ta^n) \in I$, so $ta \in \sqrt{I}$ and $a/s = (ta)/(ts) \in S^{-1}\sqrt{I}$. This proves the reverse inclusion. □

Corollary 7.19 (Nilradicals localize; AM 3.12)**.** *If $\mathfrak{n}$ is the nilradical of A, the nilradical of $S^{-1}A$ is $S^{-1}\mathfrak{n}$.*

Proof. The nilradical is $\sqrt{(0)}$. Apply the radical identity in Proposition 7.18 with $I = (0)$; extension of the zero ideal is the zero ideal. □

Corollary 7.20 (Primes in a local ring of fractions; AM 3.13)**.** *The prime ideals of $A_\mathfrak{p}$ correspond bijectively and in an inclusion-preserving way to the prime ideals of A contained in $\mathfrak{p}$.*

Proof. Take $S = A \setminus \mathfrak{p}$ in Proposition 7.18(4). A prime $\mathfrak{q}$ is disjoint from this S exactly when $\mathfrak{q} \subseteq \mathfrak{p}$. Extension and contraction preserve inclusion by their definitions. □

Example 7.21. In $A = k[x,y]$, localization at (x,y) keeps $(0),(x),(y),(x,y)$, among many other primes contained in (x,y), but discards $(x-1,y)$ because $x-1$ is an allowed denominator. Quotienting by (x), in contrast, keeps the primes containing (x). Thus localization looks downward in the prime-inclusion order, whereas a quotient looks upward.

Proposition 7.22 (Annihilators localize for finite modules; AM 3.14)**.** *If M is finitely generated, then $S^{-1}\operatorname{Ann}_A(M) = \operatorname{Ann}_{S^{-1}A}(S^{-1}M)$.*

Proof. A fraction with numerator annihilating M annihilates every fraction of M, proving one inclusion. For the other, suppose a/s annihilates $S^{-1}M$, and choose generators $m_1,\dots,m_r$. For each i, the equality $am_i/s = 0$ supplies $t_i \in S$ with $t_i a m_i = 0$. Put $t = \prod_i t_i$. Then ta kills every generator and hence all of M. Thus $ta \in \operatorname{Ann}_A(M)$ and $a/s = (ta)/(ts)$ lies in its localization. For the zero module use the empty product $t = 1$. □

Corollary 7.23 (Colon ideals localize under a finite-generation hypothesis; AM 3.15)**.** *If $N, P \subseteq M$ and P is finitely generated, then*

$$S^{-1}(N : P) = (S^{-1}N : S^{-1}P).$$

Proof. The module $(N+P)/N$ is a quotient of the finitely generated module P, so it is finitely generated. Its annihilator is $(N : P)$ by Proposition 3.15. Apply Proposition 7.22, then identify its localization with $(S^{-1}N + S^{-1}P)/S^{-1}N$ using Corollary 7.10. Its annihilator is the colon ideal on the right, again by Proposition 3.15. □

Proposition 7.24 (When a prime is the contraction of a prime; AM 3.16)**.** *Let $f : A \to B$ be a ring homomorphism and $\mathfrak{p} \in \operatorname{Spec} A$. Then there exists $\mathfrak{q} \in \operatorname{Spec} B$ with $f^{-1}(\mathfrak{q}) = \mathfrak{p}$ if and only if $\mathfrak{p}^{ec} = \mathfrak{p}$, where extension and contraction now refer to f.*

Proof. If $\mathfrak{p} = f^{-1}(\mathfrak{q})$, then $\mathfrak{p}^e \subseteq \mathfrak{q}$, so $\mathfrak{p}^{ec} \subseteq f^{-1}(\mathfrak{q}) = \mathfrak{p}$. The opposite inclusion always holds, giving equality.

Conversely assume $\mathfrak{p}^{ec} = \mathfrak{p}$, and let $T = f(A \setminus \mathfrak{p}) \subseteq B$. The ideal $\mathfrak{p}^e$ does not meet T: if $f(a) \in \mathfrak{p}^e$ with $a \notin \mathfrak{p}$, contraction would give $a \in \mathfrak{p}^{ec} = \mathfrak{p}$, a contradiction. Thus its extension to $T^{-1}B$ is proper by Proposition 7.18(2). Choose a maximal ideal $\mathfrak{n}$ containing that extension. Contract it to a prime $\mathfrak{q} \subseteq B$. This prime contains $\mathfrak{p}^e$ and avoids T. Its contraction to A therefore contains $\mathfrak{p}$ and contains no element outside $\mathfrak{p}$; hence it is exactly $\mathfrak{p}$. □

7.9 Exercises

Exercise 7.1. Let A be a ring and let $f \in A$ be nilpotent. Determine A_f. Also determine $(\mathbb{Z}/12\mathbb{Z})_3$, where 3 denotes the residue class of the integer 3.

Exercise 7.2. Compute $(\mathbb{Z}/12\mathbb{Z})_{(2)}$ and $(\mathbb{Z}/12\mathbb{Z})_{(5)}$, regarded as localizations of $\mathbb{Z}$-modules.

Exercise 7.3. Let $I \subseteq A$ be an ideal and $a \in A$. Prove that $a \in I$ if and only if $a/1 \in IA_{\mathfrak{m}}$ for every maximal ideal $\mathfrak{m}$.

Exercise 7.4 (Two successive choices of denominators; AM Chapter 3, Exercise 3)**.** Let S,T be multiplicatively closed subsets of A, let $ST = \{st : s \in S, t \in T\}$, and let U be the image of T in $S^{-1}A$. Prove $(ST)^{-1}A \cong U^{-1}(S^{-1}A)$.

Exercise 7.5 (Reducedness is local, but being a domain is not; AM Chapter 3, Exercise 5)**.** If every $A_{\mathfrak{p}}$ has no nonzero nilpotent elements, prove the same for A. If every $A_{\mathfrak{p}}$ is an integral domain, must A be an integral domain?

Exercise 7.6 (Torsion and fractions; AM Chapter 3, Exercise 12)**.** Let A be an integral domain, $K = \mathrm{Frac}(A)$, and let $T(M)$ be the torsion submodule of an A-module M. Prove that $M/T(M)$ is torsion-free; maps $f : M \to N$ send $T(M)$ into $T(N)$; the operation T preserves exactness of sequences $0 \to M' \to M \to M''$ at the first two terms; and

$$T(M) = \ker(M \longrightarrow K \otimes_A M), \qquad m \longmapsto 1 \otimes m.$$

Exercise 7.7 (Torsion-free is a local property; AM Chapter 3, Exercise 13)**.** Let A be a domain, let M be an A-module, and let $S \subseteq A$ be multiplicatively closed with $0 \notin S$. Prove $T(S^{-1}M) = S^{-1}T(M)$, and deduce that M is torsion-free if and only if all $M_{\mathfrak{p}}$ are torsion-free, if and only if all $M_{\mathfrak{m}}$ are torsion-free.

Lecture 8
Local rings and Nakayama's lemma

For a finitely generated module over a local ring, generators can be detected in a vector space over the residue field. Nakayama's lemma proves this assertion using a finite determinant calculation. Reading: [1, Chapter 2, the discussion of generators and Nakayama's lemma], together with the localization examples in Chapter 3.

8.1 Local rings and residue fields

Recall Definition 7.7: a local ring has a unique maximal ideal $\mathfrak{m}$, and its residue field is $\kappa = A/\mathfrak{m}$. The notation $(A, \mathfrak{m}, \kappa)$ records all three objects.

The symbol κ is deliberate: the residue field need not equal a ground field originally present in the problem. For example, $\mathbb{Z}_{(p)}$ has residue field $\mathbf{F}_p$; $\mathbb{R}[x]_{(x^2+1)}$ has residue field $\mathbb{C}$.

Proposition 8.1. *In a local ring $(A, \mathfrak{m})$, an element is a unit if and only if it does not belong to $\mathfrak{m}$. Therefore $1 - a$ is a unit whenever $a \in \mathfrak{m}$.*

Proof. A unit is in no proper ideal. If b is not a unit, the principal ideal (b) is proper and lies in a maximal ideal, necessarily $\mathfrak{m}$. Finally, if both a and $1 - a$ belonged to $\mathfrak{m}$, their sum 1 would also belong to it. Hence $1 - a \notin \mathfrak{m}$ and is a unit. □

A homomorphism $f : (A, \mathfrak{m}) \to (B, \mathfrak{n})$ of local rings is called *local* if $f(\mathfrak{m}) \subseteq \mathfrak{n}$. In that case it induces a map of residue fields

$$A/\mathfrak{m} \longrightarrow B/\mathfrak{n}, \qquad a + \mathfrak{m} \longmapsto f(a) + \mathfrak{n}.$$

The condition makes this formula well defined. The map is unital, hence injective, because a unital map from a field to a nonzero field has zero kernel. Also $f^{-1}(\mathfrak{n}) = \mathfrak{m}$: an element outside $\mathfrak{m}$ is a unit, so its image cannot lie in $\mathfrak{n}$.

For $A = k[x]_{(x-a)}$, a fraction f/g has $g(a) \neq 0$. It is a unit precisely when $f(a) \neq 0$. Evaluation gives

$$A \longrightarrow k, \qquad f/g \longmapsto f(a)/g(a),$$

whose kernel is $(x-a)A$. In this example "reduce modulo the maximal ideal" means "evaluate at the chosen point." In two variables, $k[x,y]_{(x,y)}$ consists of fractions whose denominator has nonzero constant term; reducing modulo (x,y) keeps only the value at the origin.

8.2 Reduction modulo the maximal ideal

Let M be an A-module. The quotient $M/\mathfrak{m}M$ is naturally a vector space over κ: define

$$(a+\mathfrak{m})(m+\mathfrak{m}M) = am+\mathfrak{m}M.$$

If a changes by an element of $\mathfrak{m}$, the result changes by an element of $\mathfrak{m}M$, so the definition makes sense. If M has r generators, their images span this vector space, and therefore its dimension is at most r.

The surprising converse is that, for finitely generated M, a spanning set after reduction lifts to generators before reduction. One might be tempted to argue by repeatedly correcting an error in $\mathfrak{m}M$, then an error in $\mathfrak{m}^2M$, and so on. That would produce an infinite process; an ordinary local ring need not permit taking its limit. Nakayama's lemma gives a finite argument.

8.2.1 Matrices over a commutative ring

The determinant and adjugate identities use addition and multiplication only, so they remain valid over any commutative ring R. To recall the notation, let $C=(c_{ij})$ be an $r\times r$ matrix, with $r\geq 1$. The identity matrix is $I_r=(\delta_{ij})$, where $\delta_{ij}=1$ for $i=j$ and 0 otherwise. Its determinant is computed by

$$\det C = \sum_{\sigma\in S_r} \operatorname{sgn}(\sigma)\prod_{i=1}^{r} c_{i,\sigma(i)}.$$

Here S_r is the set of permutations of $\{1,\dots,r\}$, and $\operatorname{sgn}(\sigma)=(-1)^e$, where e is the number of pairs $i<j$ with $\sigma(i)>\sigma(j)$. The *cofactor* of c_{ij} is $(-1)^{i+j}$ times the determinant of the matrix obtained by deleting row i and column j. The *adjugate* $\operatorname{adj}(C)$ is the transpose of the matrix of cofactors. For $r=1$ the relevant empty determinant is 1.

Expanding a determinant along a row or column gives

$$\operatorname{adj}(C)C = C\operatorname{adj}(C) = (\det C)I_r.$$

Indeed, a diagonal entry of either product is the cofactor expansion of $\det C$. An off-diagonal entry is the corresponding expansion of a determinant with two equal rows or columns; pairing the terms that interchange those rows or columns shows that it is zero. For example,

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \operatorname{adj}(C) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \qquad \det C = ad - bc.$$

In particular, when $\det C$ is a unit, $(\det C)^{-1}\operatorname{adj}(C)$ is an inverse matrix.

The module $\operatorname{End}_A(M)$ has a multiplication given by composition: $\alpha\beta = \alpha \circ \beta$, with identity id_M. This ring need not be commutative. Scalar multiplication maps do commute with every A-linear endomorphism, however. Write $\varphi^0 = \mathrm{id}_M$ and let φ^j be the j-fold composite for $j > 0$. For one fixed endomorphism φ, expressions $a_0\,\mathrm{id}_M + a_1\varphi + \cdots + a_n\varphi^n$ therefore form a commutative subring. This is the ring in which we will use the determinant identity. Saying that a polynomial $p(T) \in A[T]$ annihilates φ means that the substitution $p(\varphi)$ is the zero endomorphism, so it sends every element of M to zero.

Proposition 8.2 (The determinant trick for an endomorphism; AM 2.4)**.** *Let M be a finitely generated A-module, I an ideal, and $\varphi : M \to M$ an A-linear endomorphism satisfying $\varphi(M) \subseteq IM$. Then φ satisfies a monic polynomial relation*

$$\varphi^r + a_1\varphi^{r-1} + \cdots + a_r\,\mathrm{id}_M = 0, \qquad a_j \in I.$$

One can in fact choose $a_j \in I^j$.

Proof. For $M \neq 0$, choose generators $m_1, \ldots, m_r$ and write

$$\varphi(m_i) = \sum_j b_{ij} m_j, \qquad b_{ij} \in I, \qquad B = (b_{ij}).$$

Let $\mathbf{m} = (m_1, \ldots, m_r)^{\mathsf{T}}$ and set $C = (\delta_{ij}\varphi - b_{ij}\,\mathrm{id}_M)$. Its entries lie in the commutative subring generated by φ and the scalar maps. Therefore

$$C\mathbf{m} = 0 \implies \operatorname{adj}(C)C\mathbf{m} = 0 \implies (\det C)\mathbf{m} = 0.$$

Since the m_i generate M, the endomorphism $\det C$ is zero. Expand

$$p(T) = \det(TI_r - B) = T^r + a_1T^{r-1} + \cdots + a_r.$$

Every term contributing to a_j contains j entries of B, so $a_j \in I^j$. Substitution gives $p(\varphi) = \det C = 0$. For $M = 0$, use $p(T) = T$. □

For a matrix B, the polynomial $\det(TI_r - B)$ is called its *characteristic polynomial*. For a free module, with $I = A$ and generators chosen as a basis, the preceding proof is the characteristic-polynomial argument. The extra point is that the generators of a general module need not be independent. In Lecture 10, multiplication by an element of an algebra will play the role of φ, producing a monic polynomial equation for that element.

Corollary 8.3 (The determinant argument for generators; AM 2.5)**.** *Let M be generated by $m_1, \ldots, m_r$, and let $I \subseteq A$ be an ideal such that $IM = M$. Then some element $1 - c$ with $c \in I$ annihilates M.*

Proof. If $r = 0$, then $M = 0$ and there is nothing to prove. For $r > 0$, write

$$m_i = \sum_{j=1}^{r} a_{ij} m_j, \qquad a_{ij} \in I.$$

Put $B = (a_{ij})$ and let $\mathbf{m}$ be the column with entries m_i. These relations say $(I_r - B)\mathbf{m} = 0$. The adjugate identity for matrices over a commutative ring is

$$\operatorname{adj}(I_r - B)(I_r - B) = \det(I_r - B) I_r.$$

It follows that $\det(I_r - B) m_i = 0$ for every generator. The determinant has the form $1 - c$ with $c \in I$: on reducing every matrix entry modulo I, the matrix becomes the identity, so its determinant becomes 1. Hence $(1-c)M = 0$.

No basis for M was used. The m_i can satisfy arbitrary relations; the matrix acts on a column of module elements. This is why the argument applies to modules that are not free. □

Proposition 8.4 (Nakayama's lemma; AM 2.6)**.** *Let M be a finitely generated A-module and let I be an ideal contained in the Jacobson radical $J(A)$, the intersection of all maximal ideals of A. If $IM = M$, then $M = 0$.*

Proof. Corollary 8.3 gives $c \in I$ with $(1-c)M = 0$. If $1 - c$ were a nonunit, some maximal ideal $\mathfrak{m}$ would satisfy

$$1 - c \in \mathfrak{m}, \qquad c \in I \subseteq J(A) \subseteq \mathfrak{m},$$

forcing $1 \in \mathfrak{m}$. Hence $1 - c$ is a unit and

$$M = (1-c)^{-1}(1-c)M = 0.$$

□

For a second proof, suppose $M \neq 0$ and choose a generating list of minimum length $r > 0$. The equality $IM = M$ gives

$$m_r = \sum_{i=1}^{r} a_i m_i, \quad a_i \in I \quad \Longrightarrow \quad m_r = (1 - a_r)^{-1} \sum_{i<r} a_i m_i.$$

Here $1 - a_r$ is a unit by the same argument. Thus $m_1, \ldots, m_{r-1}$ generate M, contradicting minimality.

Corollary 8.5 (Nakayama for a quotient; AM 2.7)**.** *Let M be finitely generated, let $N \subseteq M$, and let $I \subseteq J(A)$. If $M = N + IM$, then $M = N$.*

Proof. The quotient $Q = M/N$ is finitely generated and satisfies $IQ = (IM + N)/N = Q$. Proposition 8.4 gives $Q = 0$, which means $M = N$. □

Theorem 8.6 (Nakayama's lemma, local form)**.** *Let $(A, \mathfrak{m})$ be local and let M be finitely generated. If $\mathfrak{m}M = M$, then $M = 0$. More generally, if $N \subseteq M$ and $M = N + \mathfrak{m}M$, then $M = N$.*

Proof. The unique maximal ideal is the Jacobson radical, so these are precisely Proposition 8.4 and Corollary 8.5 with $I = \mathfrak{m}$. □

The general Jacobson-radical form is useful even when A has several maximal ideals; the local form is especially concrete because its quotient is a field.

8.3 Minimal generators

For a finitely generated module M, denote by $\mu_A(M)$ the smallest number of elements in a generating set; put $\mu_A(0) = 0$. Such a minimum exists because at least one finite generating set exists. The next result computes this integer over a local ring.

Proposition 8.7 (Lifting a basis of the residue module; AM 2.8)**.** *For a finitely generated module M over $(A, \mathfrak{m}, \kappa)$, elements $m_1, \dots, m_r$ generate M if and only if their images span $M/\mathfrak{m}M$. The minimum possible number of generators is*

$$\mu_A(M) = \dim_\kappa(M/\mathfrak{m}M).$$

Proof. Put $N = \sum_{i=1}^{r} Am_i$, and let $\bar{m}_i = m_i + \mathfrak{m}M$. Then

$$\sum_i \kappa \bar{m}_i = (N + \mathfrak{m}M)/\mathfrak{m}M.$$

Hence

$$\sum_i \kappa \bar{m}_i = M/\mathfrak{m}M \iff M = N + \mathfrak{m}M \iff M = N,$$

where the last equivalence is Nakayama's lemma. If $d = \dim_\kappa(M/\mathfrak{m}M)$, every generating set has at least d elements, while lifts of a d-element basis generate M. Thus

$$\mu_A(M) \geq d, \qquad \mu_A(M) \leq d,$$

proving the formula. □

For a nonzero polynomial in $k[x,y]$, its *order at the origin* is the smallest total degree of a monomial with nonzero coefficient. For example, $x^2 + xy + y^3$ has order 2. The order of a product is the sum of the orders: multiply the terms of smallest total degree in the two factors; their product is nonzero because $k[x,y]$ is a domain, and all other terms have higher degree. A denominator with nonzero constant term has order 0. Thus a fraction f/g in $k[x,y]_{(x,y)}$ has order equal to the order of f. This is independent of the chosen fraction, since $f/g = f'/g'$ implies $fg' = f'g$ in the domain. We assign order $+\infty$ to zero.

Example 8.8. In $A = k[x,y]_{(x,y)}$, let $I = (x^2, xy, y^2)$. The quotient $I/\mathfrak{m}I$ has basis given by the three displayed quadratic monomials. They span because they generate I. They are independent because $\mathfrak{m}I$ consists of terms of order at least 3, whereas a nonzero constant linear combination of those monomials has order 2. A denominator with nonzero constant term does not change the lowest nonzero degree. Consequently I requires three generators, although it lies in a ring described by two variables.

For a module M over a domain A, its *rank* is

$$\operatorname{rank}_A M = \dim_K(K \otimes_A M), \qquad K = \operatorname{Frac}(A).$$

For a free module A^r, this agrees with the basis definition because $K \otimes_A A^r \cong K^r$. A nonzero ideal $I \subseteq A$ has rank one: localization embeds $K \otimes_A I$ into K, and any nonzero $a \in I$ gives $1 = a/a$ in the localized ideal, so that ideal is all of K. The example above therefore has rank one and needs three generators. Rank one does not imply that an ideal is principal.

8.4 Consequences and hypotheses of Nakayama's lemma

Suppose $\varphi : M \to N$ is a homomorphism of modules over a local ring, with N finitely generated. If its reduction

$$M/\mathfrak{m}M \longrightarrow N/\mathfrak{m}N$$

is surjective, then $N = \operatorname{im}\varphi + \mathfrak{m}N$, so Nakayama says φ is surjective. There is no corresponding injectivity test from the residue-field map alone: the quotient map $A \to A/(x)$ over $k[x]_{(x)}$ reduces to the identity on k but has kernel (x).

For a square matrix $C \in M_r(A)$, however, invertibility *is* detected modulo $\mathfrak{m}$. If its reduction is invertible, then $\det C$ has nonzero residue and is a unit in A. The adjugate formula gives $C^{-1} = (\det C)^{-1}\operatorname{adj}(C)$.

Caution 8.9. Both the finite-generation assumption and the condition on the ideal matter. Let $A = \mathbb{Z}_{(p)}$ and view $\mathbb{Q}$ as an A-module. Then $p\mathbb{Q} = \mathbb{Q}$, but $\mathbb{Q} \neq 0$. The module is not finitely generated: finitely many rational generators have bounded powers of p in their denominators, whereas $\mathbb{Q}$ contains arbitrarily large such powers. For the other assumption, take $A = \mathbb{Z}$, $M = \mathbb{Z}/2\mathbb{Z}$, and $I = (3)$. Then $IM = M \neq 0$. Here I is not contained in the Jacobson radical of $\mathbb{Z}$.

8.5 Locality and finite generation

Caution 8.10. Locality and finite generation have real content. Over $\mathbb{Z}$, the two nonzero finite modules $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ have zero tensor product. Over the local ring $\mathbb{Z}_{(p)}$, the nonzero modules $\mathbb{Q}$ and $\mathbf{F}_p$ also have zero tensor product: $\mathbb{Q} \otimes_A A/(p) \cong \mathbb{Q}/p\mathbb{Q} = 0$. Here $\mathbb{Q}$ is not finitely generated.

Nonzero finite modules over a local ring have nonzero tensor product

Let $(A, \mathfrak{m}, \kappa)$ be local and let M, N be nonzero finitely generated modules. Nakayama gives $M/\mathfrak{m}M \neq 0$ and $N/\mathfrak{m}N \neq 0$. Base change and the quotient tensor formula identify

$$(M \otimes_A N) \otimes_A \kappa \cong (M/\mathfrak{m}M) \otimes_\kappa (N/\mathfrak{m}N).$$

The right side is nonzero. To verify this without merely counting dimensions, take nonzero vectors v, w, choose linear functionals λ, μ with $\lambda(v) = \mu(w) = 1$, and apply the induced map $\lambda \otimes \mu$; it sends $v \otimes w$ to 1. If $M \otimes_A N$ were zero, its base change would be zero, a contradiction. This proof pinpoints the finite-generation hypothesis: it is used to make the residue vector spaces nonzero by Nakayama.

Lifting generators does not prove their independence

Over $A = k[t]_{(t)}$, the module $M = A/(t^2)$ has $M/tM \cong k$. Lifting its basis gives a generator $\bar{1}$ of M, but the map $A \to M$, $a \mapsto a\bar{1}$, has nonzero kernel (t^2). Thus $\dim_k M/tM = 1$ measures the minimal number of generators, not a free rank.

There is a useful strengthening when P is finitely generated and projective over A. Choose lifts of a basis of $P/\mathfrak{m}P$, of size r. Nakayama gives a surjection $A^r \to P$, and projectivity splits it:

$$A^r \cong P \oplus K.$$

The direct summand K is finitely generated: project the r standard basis vectors onto it. Reduction modulo $\mathfrak{m}$ gives

$$\kappa^r \cong (P/\mathfrak{m}P) \oplus (K/\mathfrak{m}K).$$

The first summand already has dimension r, so $K/\mathfrak{m}K = 0$. Nakayama forces $K = 0$. Hence every finite projective module over a local ring is free. The splitting is the extra information missing in the example $A/(t^2)$.

8.6 Support of a module

A module can carry information at some prime ideals and disappear at others. Its *support* is the set

$$\operatorname{Supp}_A(M) = \{\mathfrak{p} \in \operatorname{Spec} A : M_\mathfrak{p} \neq 0\}.$$

For an ideal I, write $V(I) = \{\mathfrak{p} \in \operatorname{Spec} A : I \subseteq \mathfrak{p}\}$. For now, $V(I)$ is simply the set of prime ideals containing I. The geometry of these sets will be studied in Lecture 16.

Example 8.11 (Axes meet at the origin)**.** For $A = k[x,y]$, let $M = A/(x)$ and $N = A/(y)$. Their supports are the two coordinate axes in $\operatorname{Spec} A$. Their tensor product is

$$A/(x) \otimes_A A/(y) \cong A/(x,y) \cong k,$$

whose support is the single maximal ideal (x,y). This is a first algebraic calculation of an intersection: quotienting by both equations retains the point where they meet. On the other hand, $A/(x) \otimes_A A/(x-1) = 0$, because the two ideals generate A. The parallel lines $x = 0$ and $x = 1$ have no intersection.

Caution 8.12. The finite-generation assumption in the formula $\operatorname{Supp}_A(M) = V(\operatorname{Ann}_A(M))$ cannot be discarded. As a $\mathbb{Z}$-module, $\mathbb{Q}/\mathbb{Z}$ has annihilator zero: for a nonzero integer n, the class of $1/(2n)$ is not killed by n. Its localization at (0) is zero, since every element is torsion. But its localization at each (p) is nonzero: the class of $1/p$ cannot be killed by any integer prime to p. Thus its support is the set of all nonzero prime ideals (p), whereas $V(\operatorname{Ann}(\mathbb{Q}/\mathbb{Z})) = \operatorname{Spec}\mathbb{Z}$ also includes (0). In fact this support cannot equal $V(I)$ for any ideal I. If $V(I)$ contains every (p), then every element of I is divisible by all prime numbers and must be zero. Hence $I = (0)$ and $V(I) = \operatorname{Spec}\mathbb{Z}$, which also contains (0).

Support identities with their hypotheses

For a finite module M,

$$\operatorname{Supp}_A(M) = V(\operatorname{Ann}_A M).$$

If $\operatorname{Ann}_A M$ contains $s \notin \mathfrak{p}$, then s is an invertible scalar on $M_\mathfrak{p}$ and also annihilates it, so $M_\mathfrak{p} = 0$. Conversely, if $M_\mathfrak{p} = 0$, choose generators $m_1, \dots, m_r$. For each i there is $s_i \notin \mathfrak{p}$ with $s_i m_i = 0$. Their product $s = \prod_i s_i$ kills every generator. Thus $s \in \operatorname{Ann}_A M \setminus \mathfrak{p}$, proving the converse. The finite product is the precise place where finiteness enters.

If $0 \to L \to M \to N \to 0$ is exact, no finiteness is required for

$$\operatorname{Supp} M = \operatorname{Supp} L \cup \operatorname{Supp} N.$$

Localize at $\mathfrak{p}$. The middle module vanishes exactly when both outside modules vanish, by exactness. For finite M, N, one also has

$$\operatorname{Supp}(M \otimes_A N) = \operatorname{Supp} M \cap \operatorname{Supp} N.$$

One containment follows because tensoring with a zero module gives zero. For the other, localize at a point in both supports and apply the nonvanishing result above over the local ring $A_\mathfrak{p}$. The finite-generation assumptions should travel with this equality whenever it is used.

8.7 Exercises

Exercise 8.1. Let $A = k[\varepsilon]/(\varepsilon^3)$. Determine all units and the unique maximal ideal.

Exercise 8.2. Let $M = A/(x^2) \oplus A/(x,y)$ for $A = k[x,y]_{(x,y)}$. Find the minimum number of generators. Does it follow that $M \cong A^2$?

Exercise 8.3. Let $(A, \mathfrak{m})$ be a local ring and M a finitely generated A-module. Suppose $m_1, \dots, m_r$ is a generating set of smallest size. Show that every relation $\sum a_i m_i = 0$ has all coefficients $a_i \in \mathfrak{m}$.

Exercise 8.4. Let $(A,\mathfrak{m})$ be a local ring, and suppose I is a finitely generated ideal of A satisfying $I^2 = I \subseteq \mathfrak{m}$. Prove that $I = 0$.

Exercise 8.5 (Surjectivity after reduction; AM Chapter 2, Exercise 10)**.** Let A be a ring, $I \subseteq J(A)$ an ideal, and $u : M \to N$ a homomorphism of A-modules. Assume N is finitely generated; M need not be finitely generated. If the induced map $M/IM \to N/IN$ is onto, prove that u is onto.

Exercise 8.6 (Nonzero finite modules over a local ring have nonzero tensor product; AM Chapter 2, Exercise 3)**.** Let $(A,\mathfrak{m},k)$ be local and let M,N be finitely generated. Prove that $M \otimes_A N = 0$ implies $M = 0$ or $N = 0$.

Exercise 8.7 (A uniform annihilating denominator; AM Chapter 3, Exercise 1)**.** Let M be a finitely generated A-module and $S \subseteq A$ a multiplicatively closed set. Prove that $S^{-1}M = 0$ if and only if there is a single $s \in S$ such that $sM = 0$.

Exercise 8.8 (Moving an ideal into the Jacobson radical; AM Chapter 3, Exercise 2)**.** For an ideal $I \subseteq A$, set $S = 1 + I = \{1 + a : a \in I\}$. Prove that S is multiplicatively closed and that $S^{-1}I$ is contained in the Jacobson radical of $S^{-1}A$. Deduce Corollary 8.3 without using determinants.

Exercise 8.9 (All the basic support identities; AM Chapter 3, Exercise 19)**.** Let A be a ring, let L,M,N denote A-modules, and let I denote an ideal of A. Prove the following statements.

1. $M \neq 0$ if and only if $\operatorname{Supp}_A(M) \neq \varnothing$.
2. $\operatorname{Supp}_A(A/I) = V(I)$.
3. For a short exact sequence $0 \to L \to M \to N \to 0$, one has $\operatorname{Supp}_A(M) = \operatorname{Supp}_A(L) \cup \operatorname{Supp}_A(N)$.
4. If $M = \sum_{i \in J} M_i$ for submodules M_i, then $\operatorname{Supp}_A(M) = \bigcup_i \operatorname{Supp}_A(M_i)$.
5. If M is finitely generated, then $\operatorname{Supp}_A(M) = V(\operatorname{Ann}_A(M))$.
6. If M,N are finitely generated, then $\operatorname{Supp}_A(M \otimes_A N) = \operatorname{Supp}_A(M) \cap \operatorname{Supp}_A(N)$.
7. If M is finitely generated, then $\operatorname{Supp}_A(M/IM) = V(I + \operatorname{Ann}_A(M))$.
8. For $f : A \to B$ and finitely generated M, let $f^* : \operatorname{Spec} B \to \operatorname{Spec} A$ send $\mathfrak{q}$ to $f^{-1}(\mathfrak{q})$. Then
$$\operatorname{Supp}_B(B \otimes_A M) = (f^*)^{-1}(\operatorname{Supp}_A(M)).$$

Exercise 8.10 (Localization and support)**.** Let $A = k[x,y]/(xy)$ and let $\mathfrak{p} = (x-1,y)$, where x,y now denote their classes in A. Identify $A_\mathfrak{p}$. For $M = A/(y)$, determine the primes $\mathfrak{q}$ for which $M_\mathfrak{q} \neq 0$.

Exercise 8.11 (A torsion-free module which is not free)**.** Let $R = k[x,y]_{(x,y)}$, with maximal ideal $\mathfrak{m} = (x,y)R$. Show that $\mathfrak{m}$ needs two generators as an R-module and is not free.

Lecture 9
Noetherian and Artinian rings

Conditions on chains of submodules express two useful kinds of finiteness. The ascending-chain condition leads to Hilbert's basis theorem: finitely many polynomial equations suffice to define any algebraic set in a fixed affine space. The descending-chain condition leads to Artinian rings, whose prime spectra consist of finitely many points. We begin with the module statements, then explain their consequences for rings. The remaining sections include the Artin–Tate lemma and the weak Nullstellensatz. The main reference is [1, Chapters 6–8].

9.1 Chain conditions

Think first of the ideals

$$(2) \supsetneq (4) \supsetneq (8) \supsetneq \cdots \quad \text{in } \mathbb{Z}.$$

For nonzero integers a, b,

$$(a) \subseteq (b) \quad \Longleftrightarrow \quad b \mid a.$$

Thus increasing divisibility gives decreasing ideals. The chain condition concerns inclusion of ideals.

Recall that a *partial order* on a set X is a reflexive, antisymmetric, and transitive relation $\leq$. A member m of a subset T is *maximal in* T if $m \leq t$ and $t \in T$ imply $t = m$; it need not be greater than every member of T. For submodules the order is inclusion. An increasing sequence *stabilizes*, or is *eventually constant*, if there is an index r such that $M_n = M_r$ for all $n \geq r$. Decreasing sequences stabilize in the same sense.

Proposition 9.1 (Chain and maximal conditions; AM 6.1)**.** *For a partially ordered set X, the following conditions are equivalent: every sequence $x_1 \leq x_2 \leq \cdots$ is eventually constant; every nonempty subset of X has a maximal element. Reversing the order gives the corresponding equivalence between the descending-chain condition and the existence of minimal elements in nonempty subsets.*

Proof. If a nonempty subset T has no maximal element, recursive choice gives

$$x_1 \in T, \qquad x_{n+1} \in T, \qquad x_n < x_{n+1} \quad (n \geq 1),$$

contradicting stabilization. Conversely, choose a maximal member x_r of the set of terms of an increasing sequence. Then

$$n \geq r \quad \Longrightarrow \quad x_r \leq x_n \quad \Longrightarrow \quad x_n = x_r.$$

Apply the same proof to the reversed order for the descending-chain statement. The recursive choice uses the usual choice principle. □

Definition 9.2. An A-module M is *Noetherian* if increasing chains of submodules stop, and *Artinian* if decreasing chains stop. A ring is Noetherian or Artinian when its underlying module over itself has the corresponding property. Its submodules are exactly its ideals.

Proposition 9.3 (Finite generation of all submodules; AM 6.2)**.** *An A-module is Noetherian if and only if every one of its submodules is finitely generated.*

Proof. Suppose M is Noetherian and $N \subseteq M$. The collection of finitely generated submodules of N contains 0, so Proposition 9.1 gives a maximal member N_0. For $x \in N$,

$$N_0 \subseteq N_0 + Ax \subseteq N, \qquad N_0 + Ax \text{ finitely generated.}$$

Maximality gives $N_0 + Ax = N_0$, hence $x \in N_0$. Thus $N = N_0$ is finitely generated.

Conversely, let $M_1 \subseteq M_2 \subseteq \cdots$. The union $N = \bigcup_i M_i$ is a submodule: any two elements occur together in a sufficiently late term. By assumption, write $N = Au_1 + \cdots + Au_r$. Choose n with $u_1, \ldots, u_r \in M_n$. Then

$$N = \sum_{i=1}^{r} Au_i \subseteq M_n \subseteq M_j \subseteq N \qquad (j \geq n).$$

Consequently $M_j = M_n = N$ for every $j \geq n$. □

An ideal generated by one element is *principal*. A *principal ideal domain* is an integral domain in which every ideal is principal; the abbreviation is PID. Thus a PID satisfies the finite-generation criterion just proved.

Example 9.4. Every ideal of $\mathbb{Z}$ is generated by one integer, so $\mathbb{Z}$ is Noetherian; the displayed chain shows that it is not Artinian. More generally, a principal ideal domain is Noetherian. In contrast,

$$k[x_1, x_2, \ldots] \supset (x_1) \subsetneq (x_1, x_2) \subsetneq \cdots$$

has an increasing chain that never stops. To verify strictness, send $x_1, \ldots, x_n$ to zero and leave x_{n+1} unchanged. This map kills $(x_1, \ldots, x_n)$ but not x_{n+1}.

Proposition 9.5 (Exact-sequence test; AM 6.3)**.** *In a short exact sequence $0 \to M' \to M \to M'' \to 0$, the module M is Noetherian if and only if M' and M'' are Noetherian. The identical assertion holds with Artinian in place of Noetherian.*

Proof. Identify M' with a submodule of M and write $\pi : M \to M''$ for the quotient. A chain in M' is already a chain in M; inverse image under π lifts a chain in M'' to one in M. Either chain condition therefore passes from M to both end terms.

Conversely, an increasing or decreasing chain (L_n) induces the chains

$$(L_n \cap M') \quad \text{and} \quad (\pi(L_n)).$$

Choose r after both stabilize. The following implication applies to any two terms beyond r, ordered as $L \subseteq K$:

$$\left.\begin{aligned} L \cap M' = K \cap M', \\ \pi(L) = \pi(K) \end{aligned}\right\} \quad \Longrightarrow \quad L = K.$$

Indeed, for $x \in K$, choose $y \in L$ with $\pi(y) = \pi(x)$. Then

$$x - y \in K \cap M' = L \cap M' \subseteq L, \qquad x = y + (x - y) \in L.$$

Hence every two sufficiently late terms coincide. □

Corollary 9.6 (Finite direct sums; AM 6.4)**.** *A finite direct sum of Noetherian modules is Noetherian. A finite direct sum of Artinian modules is Artinian.*

Proof. For two summands use the exact sequence $0 \to M_1 \to M_1 \oplus M_2 \to M_2 \to 0$, with inclusion and projection. Proposition 9.5 applies. Induction adds the remaining summands one at a time. □

Proposition 9.7 (Finite modules over finite-chain rings; AM 6.5)**.** *If A is Noetherian, every finitely generated A-module is Noetherian. If A is Artinian, every finitely generated A-module is Artinian.*

Proof. A generating list $u_1, \dots, u_r$ gives an exact sequence

$$0 \longrightarrow \ker \varphi \longrightarrow A^r \xrightarrow{\varphi} M \longrightarrow 0, \qquad \varphi(a_1, \dots, a_r) = \sum_i a_i u_i.$$

By Corollary 9.6, A^r has the chain condition satisfied by A. Proposition 9.5 passes it to M. □

Proposition 9.8 (Quotient rings; AM 6.6)**.** *If A is Noetherian or Artinian, so is A/I for every ideal I.*

Proof. The quotient is a cyclic A-module, so Proposition 9.7 applies. Its A-submodules and its A/I-submodules are the same sets: the scalar action of A factors through A/I. Consequently the module chain condition is exactly the required condition on ideals of A/I. □

Why finite modules over a Noetherian ring are finitely presented

Suppose A is Noetherian. We first prove by induction that every submodule of A^n is finitely generated. For $n = 1$ this is the definition for ideals. For $n > 1$, let $K \subseteq A^n$ and project onto the last coordinate. Its image is an ideal $J \subseteq A$, and its kernel is a submodule $K' \subseteq A^{n-1}$. Choose generators $a_1, \dots, a_s$ of J, lifts $v_1, \dots, v_s \in K$, and generators $w_1, \dots, w_t$ of K'. For $v \in K$, write its last coordinate as $\sum_i c_i a_i$. Then

$$v - \sum_i c_i v_i \in K'.$$

It follows that $v_1, \dots, v_s, w_1, \dots, w_t$ generate K.

Now if M is finite, choose $A^n \twoheadrightarrow M$. Its kernel K has a finite generating list by the preceding argument. Using that list as columns of a matrix gives a finite presentation

$$A^r \longrightarrow A^n \longrightarrow M \longrightarrow 0.$$

The proof does not assert that K is free or that the first arrow is injective. Over $k[x,y]$, the ideal (x,y) is a familiar finite submodule which is not free.

To see what can fail without the Noetherian hypothesis, take $A = k[x_1, x_2, \dots]$ and $I = (x_1, x_2, \dots)$. The ideal I is not finitely generated. A finite list of elements of I involves only $x_1, \dots, x_N$ for some N; setting these variables to zero kills that list but not x_{N+1}. The cyclic module A/I is not finitely presented. Indeed, if a module is finitely presented, the kernel of any surjection from a finite free module is finitely generated. Here is a justification of that last fact. Given a finite presentation with kernel $K_0 \subseteq A^s$ and another surjection $F \to M$, form

$$T = \{(u,v) \in A^s \oplus F : \pi(u) = q(v)\}.$$

Projection onto either free factor splits, giving $T \cong K_0 \oplus F$ and $T \cong \ker(q) \oplus A^s$. The first module is finite; hence the direct summand $\ker(q)$ is finite. Apply this with $F = A$ to obtain the contradiction.

9.2 Polynomial rings and Hilbert's basis theorem

Proposition 9.9 (Surjective images; AM 7.1)**.** *A surjective ring homomorphism from a Noetherian ring has Noetherian target.*

Proof. For $f : A \twoheadrightarrow B$, the first isomorphism theorem gives $B \cong A/\ker f$. Apply Proposition 9.8. This also explicitly identifies the result with the quotient-ring statement already proved. □

Proposition 9.10 (Module-finite extensions; AM 7.2)**.** *If $A \subseteq B$, A is Noetherian, and B is finitely generated as an A-module, then B is a Noetherian ring.*

Proof. Proposition 9.7 makes B a Noetherian A-module. Every ideal of B is an A-submodule, because multiplication by elements of $A \subseteq B$ preserves it. An increasing chain of ideals is therefore an increasing chain among those submodules, and it stops. □

Example 9.11. The Gaussian integers $\mathbb{Z}[i]$ are generated as a $\mathbb{Z}$-module by $1, i$, because $i^2 = -1$ reduces every polynomial expression to $a + bi$. They form a Noetherian ring by Proposition 9.10. Being a finite module is stronger than being a finitely generated algebra: $\mathbb{Z}[x]$ is generated as an algebra by x, but its powers are linearly independent over $\mathbb{Z}$.

Proposition 9.12 (Localization; AM 7.3)**.** *If A is Noetherian and S is multiplicatively closed, then $S^{-1}A$ is Noetherian.*

Proof. For an ideal $J \subseteq S^{-1}A$, let $I = \{a \in A : a/1 \in J\}$. Since $s/1$ is a unit for $s \in S$,

$$\frac{a}{s} \in J \quad \Longleftrightarrow \quad \frac{a}{1} \in J \quad \Longleftrightarrow \quad a \in I.$$

Thus $J = S^{-1}I$. Write $I = (a_1, \ldots, a_r)$ using Noetherianity. Then

$$J = S^{-1}I = \left(\frac{a_1}{1}, \ldots, \frac{a_r}{1}\right).$$

Every ideal of $S^{-1}A$ is therefore finitely generated. □

Corollary 9.13 (Local rings of a Noetherian ring; AM 7.4)**.** *For a prime ideal $\mathfrak{p}$ of a Noetherian ring A, the ring $A_{\mathfrak{p}}$ is Noetherian.*

Proof. Use Proposition 9.12 with $S = A \setminus \mathfrak{p}$, which is multiplicatively closed because $\mathfrak{p}$ is prime. □

Theorem 9.14 (Hilbert's basis theorem; AM 7.5)**.** *If A is Noetherian, then the polynomial ring $A[x]$ is Noetherian.*

Proof. Fix a nonzero ideal $I \subseteq A[x]$. Let L consist of zero and the leading coefficients of nonzero members of I. First we check that L is an ideal of A. If $a, b \in L$ are represented by f, g, multiply the lower-degree polynomial by a power of x to make their degrees equal. If $a + b \neq 0$, the sum has leading coefficient $a + b$; if $a + b = 0$, it is already in L. For $c \in A$, either $ca = 0$, or cf has leading coefficient ca. This checks both ideal properties.

Choose generators $a_1, \ldots, a_r$ of L and polynomials $f_i \in I$ with those leading coefficients. Set $d_i = \deg f_i$ and $d = \max d_i$. The polynomials of degree less than d form a finite free A-module $P_{<d}$, with basis $1, x, \ldots, x^{d-1}$; if $d = 0$, it is 0. By Propositions 9.7 and 9.3, the submodule $I \cap P_{<d}$ has finitely many module generators $g_1, \ldots, g_s$.

We claim $I = (f_1, \ldots, f_r, g_1, \ldots, g_s)$. If $f \in I$ has degree $n \geq d$, write its leading coefficient as $\sum_i c_i a_i$ and put

$$f^{(1)} = f - \sum_i c_i x^{n-d_i} f_i.$$

The degree-n coefficient vanishes, so either $f^{(1)} = 0$ or $\deg f^{(1)} < \deg f$. Repetition gives

$$f = \sum_i h_i f_i + r, \qquad r \in I \cap P_{<d} = \sum_j A g_j.$$

The process terminates because nonnegative degrees strictly decrease. For $\deg f < d$, take $r = f$ immediately. Hence the displayed finite list generates I. The zero ideal is generated by 0. □

No leading coefficient was divided by another. This is the precise feature that lets Hilbert's theorem work over $\mathbb{Z}$ or over a ring with zero divisors, where the polynomial division algorithm over a field is unavailable.

Corollary 9.15 (Finitely many variables; AM 7.6)**.** *If A is Noetherian, then $A[x_1,\dots,x_n]$ is Noetherian for every finite n.*

Proof. Induct on n, viewing the ring at the next step as $A[x_1,\dots,x_{n-1}][x_n]$ and applying Theorem 9.14. □

Corollary 9.16 (Finitely generated algebras; AM 7.7)**.** *Every finitely generated algebra over a Noetherian ring is Noetherian. In particular, every finitely generated ring and every finitely generated algebra over a field is Noetherian.*

Proof. If $B = A[b_1,\dots,b_n]$, evaluation gives a surjection $A[x_1,\dots,x_n] \twoheadrightarrow B$. Apply Corollary 9.15 and Proposition 9.9. A finitely generated ring means a finitely generated $\mathbb{Z}$-algebra here; $\mathbb{Z}$ and every field are Noetherian. □

Example 9.17 (Finitely many equations)**.** For a family $f_\lambda \in k[x_1,\dots,x_n]$, Hilbert's basis theorem gives

$$I = (f_\lambda : \lambda \in \Lambda) = (g_1,\dots,g_r).$$

Each g_i is a finite sum of multiples of the f_λ. Combining the finitely many indices occurring in these sums gives a finite set $F \subseteq \Lambda$ with

$$(g_1,\dots,g_r) \subseteq (f_\lambda : \lambda \in F) \subseteq I.$$

Both inclusions are equalities. Therefore

$$\{a \in k^n : f_\lambda(a) = 0\ (\lambda \in \Lambda)\} = \{a \in k^n : f_\lambda(a) = 0\ (\lambda \in F)\}.$$

Only finitely many of the original equations are needed.

9.3 Finite algebras and the weak Nullstellensatz

The next result detects finite generation of an intermediate ring. This is useful because subrings of Noetherian rings need not be Noetherian. For example, the polynomial ring in infinitely many variables embeds in its fraction field, and every field is Noetherian.

For an inclusion of rings $B \subseteq C$, an element $c \in C$ is *integral over B* if it satisfies a monic equation $c^d + b_{d-1}c^{d-1} + \cdots + b_0 = 0$, with $b_i \in B$. Such an equation expresses all powers c^n with $n \geq d$ as B-linear combinations of $1, c, \dots, c^{d-1}$. Consequently, if $c_1,\dots,c_s$ are integral, the finitely many monomials $c_1^{e_1}\cdots c_s^{e_s}$ with $0 \leq e_i < d_i$ span $B[c_1,\dots,c_s]$ as a B-module. The extension $B \subseteq C$ is called *integral* if every element of C is integral over B.

The preceding calculation will suffice in this section; Lecture 10 develops the remaining properties.

Proposition 9.18 (Artin–Tate lemma; AM 7.8)**.** *Let $A \subseteq B \subseteq C$ be rings. Suppose A is Noetherian, C is a finitely generated A-algebra, and either C is a finitely generated B-module or C is integral over B. Then B is a finitely generated A-algebra.*

Proof. Write $C = A[x_1, \dots, x_m]$. Since $A \subseteq B \subseteq C$, we also have $C = B[x_1, \dots, x_m]$. If C is integral over B, the preceding monomial calculation makes it a finite B-module. It is therefore enough to handle that case.

Choose B-module generators $y_1, \dots, y_n$ of C, including $y_1 = 1$. Because they span C, we can write

$$x_i = \sum_j b_{ij} y_j, \qquad y_j y_k = \sum_\ell c_{jk\ell} y_\ell, \qquad b_{ij}, c_{jk\ell} \in B.$$

Let B_0 be the A-subalgebra generated by this finite list of coefficients. Hilbert's basis theorem makes B_0 Noetherian. The B_0-span $D = \sum_j B_0 y_j$ contains 1 and every x_i. It is closed under multiplication by the displayed product formulas. Thus D is a subring of C containing $A[x_1, \dots, x_m] = C$, so $D = C$.

We have proved that C is a finite B_0-module. It is therefore a Noetherian B_0-module. Its submodule B has finite B_0-module generators $z_1, \dots, z_t$. Adjoin to A the generators of B_0 and the z_i. Every element of B is a B_0-linear combination of the z_i, so this finite list generates B as an A-algebra. □

9.3.1 Field extensions and polynomial relations

Definition 9.19. A *field extension* E/k is an inclusion of fields $k \subseteq E$; the slash does not denote a quotient. Multiplication by elements of k makes E a k-vector space. Its dimension is the *degree* $[E : k]$, and the extension is *finite* if this dimension is finite. An element $\alpha \in E$ is *algebraic over* k if $f(\alpha) = 0$ for some nonzero $f \in k[T]$, and *transcendental over* k otherwise. The extension is *algebraic* if every element of E is algebraic over k.

The notation $k[\alpha]$ means the subring consisting of polynomial expressions in α. The notation $k(\alpha)$ means the smallest subfield containing k and α, so it also permits divisions by nonzero expressions. For example, the polynomial variable t is transcendental over k, and $k[t]$ is strictly smaller than $k(t)$ because $1/t \notin k[t]$. Algebraic elements behave differently.

In a domain, two nonzero elements are *associates* if one is a unit times the other. A nonzero nonunit is *irreducible* if every factorization of it has a unit factor. In $k[T]$, the units are the nonzero constants, since degrees add under multiplication. Thus a positive-degree polynomial is irreducible exactly when it cannot be written as a product of two polynomials of positive degree.

Proposition 9.20 (A single algebraic generator)**.** *If α is algebraic over k, there is a unique monic polynomial $m_\alpha \in k[T]$ of smallest positive degree with $m_\alpha(\alpha) = 0$. It is irreducible, every polynomial vanishing at α is divisible by it, and*

$$k[\alpha] = k(\alpha) \cong k[T]/(m_\alpha).$$

If $d = \deg m_\alpha$, then $1, \alpha, \ldots, \alpha^{d-1}$ is a k-basis of this field. The polynomial m_α is called the minimal polynomial *of α over k.*

Proof. Choose a nonzero polynomial of smallest degree vanishing at α and divide it by its leading coefficient. Its degree is positive because a nonzero constant cannot vanish in a field. If it factored into two polynomials of positive degree, their values at α would have product zero, so one factor would vanish, contrary to minimality. Hence it is irreducible.

Division with remainder writes any polynomial as $qm_\alpha + r$, where $\deg r < d$ or $r = 0$. If the polynomial vanishes at α, then so does r, and minimality forces $r = 0$. This proves both the divisibility assertion and uniqueness. Evaluation and the first isomorphism theorem now identify $k[\alpha]$ with $k[T]/(m_\alpha)$.

This quotient is a field. Indeed, if f is not divisible by m_α, the ideal (f, m_α) in the principal ideal domain $k[T]$ is the whole ring: a generator divides the irreducible polynomial m_α and cannot be associated to it. Thus $uf + vm_\alpha = 1$ for some $u, v \in k[T]$, which supplies an inverse for the class of f. Finally, division gives the stated spanning list, and a dependence among its elements would be a nonzero polynomial of degree less than d vanishing at α. Thus the list is a basis. □

A finite field extension is algebraic: in an n-dimensional space, the $n+1$ elements $1, \alpha, \ldots, \alpha^n$ are dependent. Conversely, adjoining finitely many algebraic elements gives a finite extension. To see this, apply the preceding proposition successively. If $k \subseteq F \subseteq E$, a basis $u_1, \ldots, u_r$ of F/k and a basis $v_1, \ldots, v_s$ of E/F give the k-basis $u_i v_j$. Spanning follows by two successive expansions; independence follows by first grouping the coefficients of each v_j, then using independence of the u_i. Consequently

$$[E : k] = [E : F][F : k]$$

for finite extensions. This is the *tower formula*. It also shows that sums, products, and nonzero quotients of algebraic elements are algebraic: they belong to a field obtained by adjoining finitely many such elements. Algebraicity is also transitive. If E/F and F/k are algebraic and $\alpha \in E$, the finitely many coefficients of an equation for α generate a finite extension F_0/k. The extension $F_0(\alpha)/F_0$ is finite, so the tower formula makes α algebraic over k.

Example 9.21. The element $\sqrt{2} \in \mathbb{R}$ has minimal polynomial $T^2 - 2$ over $\mathbb{Q}$. It has no rational root: a reduced fraction a/b with $(a/b)^2 = 2$ would force both a and b to be even. A quadratic polynomial without a root cannot factor into positive-degree factors. Thus $\mathbb{Q}(\sqrt{2})$ has basis $1, \sqrt{2}$ and degree 2. For instance,

$$(1 + \sqrt{2})^{-1} = \sqrt{2} - 1$$

already lies in $\mathbb{Q}[\sqrt{2}]$.

Definition 9.22. Elements $x_1, \ldots, x_r$ in a k-algebra are *algebraically independent over k* if the evaluation map $k[T_1, \ldots, T_r] \to A$, $T_i \mapsto x_i$, is injective. Equivalently, no nonzero polynomial relation with coefficients in k holds among them. If a nonzero relation exists,

they are *algebraically dependent*. For an independent tuple in a field, $k[x_1,\dots,x_r]$ is a polynomial ring, and $k(x_1,\dots,x_r)$ is its fraction field. In a nonzero k-algebra the empty tuple is independent, and in an extension field the corresponding subring and subfield are both k.

For example, the symbols u,v are independent in $k[u,v]$. The two elements $t^2,t^3 \in k[t]$ are dependent because $(t^3)^2-(t^2)^3=0$. A single element is independent precisely when it is transcendental.

Recall that k is *algebraically closed* if every nonconstant polynomial over k has a root in k. Removing one linear factor at a time then shows that every polynomial splits into linear factors. An *algebraic closure* $\bar{k}$ is an algebraic extension of k that is algebraically closed. We will use the standard existence theorem; the following construction explains why such a field is available.

For a field F, introduce a variable X_f for each monic nonconstant polynomial $f \in F[T]$. In the polynomial ring on these variables, each polynomial still uses only finitely many variables. The ideal generated by the elements $f(X_f)$ is proper. Otherwise an expression for 1 would involve only finitely many of these generators. Adjoin roots of these finitely many polynomials successively, using quotients by irreducible factors, and evaluate all the variables occurring in that expression; it would give $1=0$ in a field. Choose a maximal ideal containing this proper ideal. The quotient is a field F' containing F in which every polynomial over F has a root. It is algebraic over F: each generating variable satisfies its chosen polynomial, and every element uses only finitely many variables.

Starting with $F_0=k$, repeat this construction to obtain $F_0 \subseteq F_1 \subseteq \cdots$. Their union is algebraic over k, by the finite-extension and tower arguments above. The coefficients of any polynomial over this union lie in some F_i, so that polynomial has a root in F_{i+1}. The union is therefore an algebraic closure. This construction uses the same maximal-ideal existence principle used earlier for rings.

9.3.2 Fields of finite algebra type

Proposition 9.23 (Zariski's lemma; AM 7.9)**.** *If a field E is a finitely generated algebra over a field k, then E is a finite algebraic extension of k.*

Proof. Write $E=k[x_1,\dots,x_n]$ and choose a largest algebraically independent subcollection of these generators. Renumber it as $x_1,\dots,x_r$. By algebraic independence, $k[x_1,\dots,x_r]$ is a polynomial ring. Maximality implies that every remaining generator is algebraic over $F=k(x_1,\dots,x_r)$: a dependence involving one new generator is a nonzero polynomial in that generator with coefficients in $k[x_1,\dots,x_r]$, which can be read over its fraction field. Adjoining finitely many algebraic elements gives a finite extension, because powers below the respective minimal-polynomial degrees span at each step. Hence E is a finite F-module.

Apply the Artin–Tate lemma to $k \subseteq F \subseteq E$. It says F is a finitely generated k-algebra. If $r>0$, write algebra generators of F as fractions $f_1/g_1,\dots,f_s/g_s$ of polynomials in the independent variables. Put $g=\prod_i g_i \neq 0$, taking $g=1$ for an empty list. Every polynomial expression in those fractions has the form f/g^N for some polynomial f and some $N \geq 0$.

Consider $h = 1 + x_1 g$. It is a nonconstant polynomial, hence not a unit in $k[x_1, \dots, x_r]$. Since $1/h \in F$, finite algebra generation would give $1/h = f/g^N$, so $g^N = hf$. But $h - x_1 g = 1$. Raising this identity to the Nth power shows that h divides 1: every term of $(h - x_1 g)^N$ contains h, including the last term because h divides g^N. For $N = 0$ the equation $1 = hf$ says this directly. This contradicts the nonconstant degree of h. Therefore $r = 0$. All generators are algebraic over k, and the finite-spanning argument above proves $[E : k] < \infty$. □

Corollary 9.24 (Weak Nullstellensatz; AM 7.10)**.** *If A is a finitely generated k-algebra and $\mathfrak{m}$ is maximal, the field $A/\mathfrak{m}$ is a finite algebraic extension of k. If k is algebraically closed, the structure map identifies $A/\mathfrak{m}$ with k.*

Proof. The quotient $A/\mathfrak{m}$ is a field of finite algebra type over k, so Proposition 9.23 gives $[A/\mathfrak{m} : k] < \infty$. If k is algebraically closed, every irreducible polynomial in $k[T]$ is linear. Thus, for $\alpha \in A/\mathfrak{m}$,

$$m_\alpha(T) = T - a \quad (a \in k) \quad \Longrightarrow \quad \alpha = a.$$

The structure map $k \to A/\mathfrak{m}$ is consequently surjective as well as injective. □

Example 9.25 (Why maximal ideals are points)**.** Let k be algebraically closed and $\mathfrak{m} \subset k[x_1, \dots, x_n]$ maximal. In the quotient $k[x_1, \dots, x_n]/\mathfrak{m} \cong k$, write $x_i \mapsto a_i$. Evaluation gives

$$(x_1 - a_1, \dots, x_n - a_n) \subseteq \mathfrak{m}, \qquad \frac{k[x_1, \dots, x_n]}{(x_1 - a_1, \dots, x_n - a_n)} \cong k.$$

The ideal on the left is maximal, hence equals $\mathfrak{m}$. In contrast,

$$\mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C}, \qquad \{a \in \mathbb{R} : a^2 + 1 = 0\} = \varnothing.$$

A maximal ideal over a field that is not algebraically closed need not correspond to a point with coordinates in that field.

9.4 Composition series and length

Definition 9.26. A nonzero module is *simple* if its only submodules are 0 and itself. A *composition series* of M is a finite chain

$$M = M_0 \supsetneq M_1 \supsetneq \cdots \supsetneq M_n = 0$$

with all M_{i-1}/M_i simple. There are n links, so its length is n. The zero module has the empty series of length 0.

Every simple module over a commutative ring is $A/\mathfrak{m}$ for a maximal ideal $\mathfrak{m}$. Indeed, a nonzero element u generates the simple module; the surjection $A \to Au$, $a \mapsto au$, has a maximal kernel because the quotient has no nontrivial ideals.

Proposition 9.27 (Length is well-defined; AM 6.7)**.** *If M has a composition series of length n, every composition series has length n, every strict chain of submodules has at most n links, and every finite chain can be refined to a composition series.*

Proof. For a module possessing a composition series, let $\ell(M)$ be the least possible length. Fix a series of this least length and a submodule N. Intersect every term with N. The natural map

$$(N \cap M_{i-1})/(N \cap M_i) \longrightarrow M_{i-1}/M_i$$

is injective, so its source is either zero or the entire simple target. Removing repeated terms gives a composition series of N, of length at most $\ell(M)$.

If $\ell(N) = \ell(M)$, none of these factors can have disappeared: otherwise we would have constructed a shorter series of N. Starting at $M_n = 0$, equality of the corresponding factors proves by upward induction that $N \cap M_i = M_i$ for every i. For the induction step, surjectivity onto M_i/M_{i+1} and the inclusion $M_{i+1} \subseteq N$ imply $M_i \subseteq N$. Hence $N = M$. We have proved that a proper submodule has strictly smaller least length.

Along a strict chain, these nonnegative integer least lengths strictly decrease. Every chain therefore has at most $\ell(M)$ links. A composition series has at least $\ell(M)$ links by definition and at most that many by this bound, so all have the same length. Finally, start with any finite chain and append M and 0 at its ends if they are absent. If this chain is not a composition series, some quotient has a proper nonzero submodule; insert its preimage. Each insertion adds one link. The bound prevents infinitely many insertions, and the resulting maximal chain is a composition series. □

The common length is denoted $\operatorname{length}_A M$, or $\ell(M)$ when the ring is fixed. We say that M has *finite length* when it has a composition series. More generally, a finite *filtration* is any nested finite list of submodules, usually starting at M and ending at 0; its quotients of successive terms are its *successive factors*. They need not be simple. A composition series is a filtration all of whose nonzero successive factors are simple.

Proposition 9.28 (Finite length and the two chain conditions; AM 6.8)**.** *A module has a composition series if and only if it is both Noetherian and Artinian.*

Proof. A composition series bounds the length of every strict chain by Proposition 9.27, so neither an increasing nor a decreasing infinite strict chain exists. Conversely, suppose both chain conditions hold. A nonzero module has a maximal proper submodule: apply the maximal-element condition to its proper submodules, which include 0. Choose such a submodule $M_1 \subsetneq M$, then a maximal proper submodule of M_1 if it is nonzero, and continue. Each quotient is simple by maximality. The descending-chain condition forces the process to reach zero after finitely many steps, giving a composition series. □

Proposition 9.29 (Additivity of length; AM 6.9)**.** *For a short exact sequence of finite-length modules $0 \to M' \to M \to M'' \to 0$,*

$$\operatorname{length}_A M = \operatorname{length}_A M' + \operatorname{length}_A M''.$$

Proof. Identify M' with its image. Take a composition series from M' down to zero. Above it, insert the inverse images in M of the terms of a composition series of M''. The inverse

image of 0 is M', so these chains join. The upper successive quotients are the simple factors of M'' by the quotient isomorphism theorem; the lower ones are those of M'. This is a composition series of M whose number of links is the sum claimed. Proposition 9.27 ensures that this count is the length, independent of the chosen series. □

Remark 9.30 (The module form of Jordan–Hölder). There is a stronger uniqueness statement: the isomorphism classes of the simple factors, counted with multiplicity, are independent of the composition series. Here is a proof rather than only a reference to the group version. Use induction on the finite length of M. Let N and N' be the penultimate terms above zero in two series; both are simple. If $N = N'$, quotient by them and apply induction. If $N \neq N'$, simplicity gives $N \cap N' = 0$, hence $(N+N')/N \cong N'$ and $(N+N')/N' \cong N$. Apply the inductive assertion in M/N to compare its given series with one passing through $(N+N')/N$, and do the same in M/N'. Above $N+N'$, choose the same series of $M/(N+N')$; its existence follows from the two chain conditions. Both original series therefore have the factors of this common quotient together with one copy each of N and N'. This proves the stronger assertion.

Proposition 9.31 (Vector spaces; AM 6.10)**.** *For a vector space V over a field k, finite dimension, finite length, the ascending-chain condition, and the descending-chain condition are equivalent. When these hold,* $\operatorname{length}_k V = \dim_k V$.

Proof. A basis $v_1, \ldots, v_d$ gives a chain obtained by successively discarding basis vectors. Its successive quotients have dimension one and are simple, so the length is d. Finite length implies both chain conditions. If V is infinite-dimensional, choose independent $v_1, v_2, \ldots$. The spans of $v_1, \ldots, v_n$ form a strictly increasing chain. The spans of $v_n, v_{n+1}, \ldots$ form a strictly decreasing chain, because v_n is not in the span of the later vectors. Thus either chain condition excludes infinite dimension. □

Corollary 9.32 (A product of maximal ideals equal to zero; AM 6.11)**.** *Suppose $\mathfrak{m}_1 \cdots \mathfrak{m}_r = 0$ for finitely many maximal ideals, with repetitions allowed. Then A is Noetherian if and only if A is Artinian.*

Proof. Filter A by $A \supseteq \mathfrak{m}_1 \supseteq \mathfrak{m}_1\mathfrak{m}_2 \supseteq \cdots \supseteq \mathfrak{m}_1 \cdots \mathfrak{m}_r = 0$. The ith factor

$$V_i = \mathfrak{m}_1 \cdots \mathfrak{m}_{i-1}/\mathfrak{m}_1 \cdots \mathfrak{m}_i$$

is annihilated by $\mathfrak{m}_i$, so is a vector space over the field $A/\mathfrak{m}_i$. Its subspaces and A-submodules coincide. Proposition 9.31 says its two chain conditions are equivalent. Repeated application of Proposition 9.5 says that either chain condition holds for A exactly when it holds for every factor. The two conditions on A are therefore equivalent. □

Example 9.33 (Length versus dimension and cardinality)**.** The $\mathbb{Z}$-module $\mathbb{Z}/12\mathbb{Z}$ has the composition series

$$\mathbb{Z}/12\mathbb{Z} \supset 2\mathbb{Z}/12\mathbb{Z} \supset 4\mathbb{Z}/12\mathbb{Z} \supset 0,$$

whose factors have orders $2, 2, 3$. Its length is 3, even though it has 12 elements. Every field, including the infinite field $\mathbb{Q}$, has length 1 over itself. The word Artinian does not mean that the underlying set is finite.

9.5 Nilpotent ideals and minimal primes

An element being nilpotent is a separate assertion for each element. An ideal being nilpotent requires one exponent that works for every product of that many ideal elements. Finite generation connects these assertions.

Proposition 9.34 (A uniform power of the radical; AM 7.14)**.** *If A is Noetherian and I is an ideal, there is an integer $N \geq 1$ such that $(\sqrt{I})^N \subseteq I$.*

Proof. Choose generators $a_1, \ldots, a_r$ of $\sqrt{I}$. For each i, choose $n_i \geq 1$ with $a_i^{n_i} \in I$. Put $N = 1 + \sum_i (n_i - 1)$. A generator of $(\sqrt{I})^N$ is a monomial $a_1^{e_1} \cdots a_r^{e_r}$ with $\sum_i e_i = N$. If every $e_i < n_i$, their sum would be at most $N-1$. Thus some $e_i \geq n_i$, and this monomial belongs to I. Since all the generating monomials belong to I, the whole power does. If $\sqrt{I} = 0$, take $N = 1$ directly. □

Corollary 9.35 (Nilradical of a Noetherian ring; AM 7.15)**.** *The nilradical of a Noetherian ring is a nilpotent ideal.*

Proof. Apply Proposition 9.34 to $I = 0$. □

Example 9.36 (Finding an exponent)**.** In $A = k[x,y]/(x^3, y^4)$, the ideal $\mathfrak{m} = (x,y)$ is the nilradical. The preceding proof gives $\mathfrak{m}^6 = 0$, since $1 + (3-1) + (4-1) = 6$. Indeed, every monomial of total degree 6 is divisible by x^3 or y^4. The bound is sharp here: the degree-five monomial x^2y^3 survives. The residue classes $x^i y^j$, $0 \leq i < 3$, $0 \leq j < 4$, form a k-basis, so the ring has vector-space dimension 12.

We need one small piece of terminology about individual ideals. A proper ideal Q is *primary* if $ab \in Q$ and $a \notin Q$ imply $b^n \in Q$ for some $n \geq 1$. Equivalently, every zero divisor in A/Q is nilpotent. If its radical is a prime $\mathfrak{p}$, one calls it $\mathfrak{p}$-primary. We will use this condition for ideals whose radical is a maximal ideal; no decomposition theory is required.

Corollary 9.37 (Ideals with maximal radical; AM 7.16)**.** *Let A be Noetherian, let $\mathfrak{m}$ be maximal, and let Q be an ideal. The following are equivalent:*

(i) *Q is $\mathfrak{m}$-primary;*

(ii) $\sqrt{Q} = \mathfrak{m}$;

(iii) $\mathfrak{m}^n \subseteq Q \subseteq \mathfrak{m}$ *for some integer $n \geq 1$.*

Proof. The first assertion includes $\sqrt{Q} = \mathfrak{m}$ in its definition. Conversely, suppose $\sqrt{Q} = \mathfrak{m}$. A maximal ideal containing Q must contain its radical and thus equal $\mathfrak{m}$. Therefore A/Q is local with unique maximal ideal $\mathfrak{m}/Q$. Every element of this maximal ideal is nilpotent, by $\sqrt{Q} = \mathfrak{m}$, and every element outside it is a unit, by the local-ring unit criterion. A zero divisor cannot be a unit, so every zero divisor is nilpotent. This proves primaryness directly.

If $\sqrt{Q} = \mathfrak{m}$, Proposition 9.34 gives $\mathfrak{m}^n \subseteq Q$ and clearly $Q \subseteq \mathfrak{m}$. Conversely, taking radicals of the inclusions in (iii) gives $\mathfrak{m} = \sqrt{\mathfrak{m}^n} \subseteq \sqrt{Q} \subseteq \sqrt{\mathfrak{m}} = \mathfrak{m}$. The identity $\sqrt{\mathfrak{m}^n} = \mathfrak{m}$ follows because an element of $\mathfrak{m}$ has its nth power in $\mathfrak{m}^n$, while a prime containing $\mathfrak{m}^n$ contains $\mathfrak{m}$. □

A prime P is *minimal over an ideal* I if $I \subseteq P$ and there is no prime Q with $I \subseteq Q \subsetneq P$. A *minimal prime of* A means a prime minimal over the zero ideal. Minimality is taken among primes, rather than among all ideals.

Lemma 9.38 (Finitely many minimal primes)**.** *For every ideal I in a Noetherian ring, there are finitely many primes minimal among those containing I. Every prime containing I contains at least one of these minimal primes.*

Proof. First, minimal primes exist below any given prime $P \supseteq I$, with no Noetherian hypothesis. Consider primes Q with $I \subseteq Q \subseteq P$, ordered by reverse inclusion. The intersection of a chain of such primes is prime: if ab lies in the intersection and a does not, choose a chain member Q_0 not containing a. Then $b \in Q_0$. For a smaller chain member Q, we still have $a \notin Q$, so $b \in Q$; for a larger member, $b \in Q_0 \subseteq Q$. Thus b is in the whole intersection. The intersection contains I and is proper because it is contained in P. Zorn's lemma gives a minimal prime below P.

For finiteness, suppose there were ideals having infinitely many minimal primes. The Noetherian maximal condition gives a maximal such ideal I. It is not prime: a prime ideal is itself its only minimal prime. Choose $a, b \notin I$ with $ab \in I$. The strictly larger ideals $I + (a)$ and $I + (b)$ each have only finitely many minimal primes. If P is minimal over I, primality forces $a \in P$ or $b \in P$. In the first case choose a prime Q minimal over $I + (a)$ contained in P, using the first paragraph. Then $I \subseteq Q \subseteq P$, so minimality over I forces $Q = P$. The second case is identical with b. Every minimal prime of I is therefore on one of two finite lists, a contradiction. □

Example 9.39 (Minimal does not mean maximal)**.** In $k[x,y]/(xy)$ the minimal primes are (x) and (y), whereas maximal ideals include $(x, y-a)$ and $(x-a, y)$ for $a \in k$ when k is algebraically closed. The two minimal primes describe the two coordinate axes. In contrast, when every prime is maximal, the lemma says that a Noetherian ring has only finitely many primes in total. This special case will be exactly what the next theorem needs.

9.6 Artinian rings and their local factors

Proposition 9.40 (Primes in an Artinian ring; AM 8.1)**.** *Every prime ideal of an Artinian ring is maximal.*

Proof. If P is prime, $B = A/P$ is both Artinian and an integral domain. For $0 \neq x \in B$, the descending chain $(x) \supseteq (x^2) \supseteq \cdots$ stabilizes. Thus $x^n = x^{n+1}y$ for some $n \geq 1$ and $y \in B$. Cancellation in the domain gives $1 = xy$. Every nonzero element is invertible, so B is a field and P is maximal. □

Corollary 9.41 (The two radicals coincide; AM 8.2)**.** *In an Artinian ring the nilradical equals the Jacobson radical.*

Proof. The nilradical is the intersection of prime ideals, and the Jacobson radical is the intersection of maximal ideals. Proposition 9.40 identifies these two collections. □

Proposition 9.42 (Only finitely many maximal ideals; AM 8.3)**.** *An Artinian ring has finitely many maximal ideals.*

Proof. The zero ring has none, so suppose $A \neq 0$. Among the finite intersections of maximal ideals, choose a minimal one $J = \mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_r$ using the descending-chain condition. For any maximal ideal $\mathfrak{m}$, the smaller intersection $J \cap \mathfrak{m}$ must equal J, so $J \subseteq \mathfrak{m}$. In particular, $\mathfrak{m}_1 \cdots \mathfrak{m}_r \subseteq \mathfrak{m}$. Primality implies $\mathfrak{m}_i \subseteq \mathfrak{m}$ for some i: otherwise choose $a_i \in \mathfrak{m}_i \setminus \mathfrak{m}$, whose product contradicts the inclusion. Since both ideals are maximal, $\mathfrak{m} = \mathfrak{m}_i$. Thus the finite list contains all maximal ideals. □

Proposition 9.43 (Nilradical of an Artinian ring; AM 8.4)**.** *The nilradical N of an Artinian ring is nilpotent.*

Proof. The chain $N \supseteq N^2 \supseteq \cdots$ stabilizes. Choose $k \geq 1$ so that $J = N^k = N^{k+1} = \cdots$. Then $J^2 = N^{2k} = J$. We claim $J = 0$.

Suppose $J \neq 0$. Among the ideals C with $JC \neq 0$, choose a minimal one by the descending-chain condition; there is at least one, namely A. Some $x \in C$ satisfies $Jx \neq 0$, for otherwise every product contributing to JC would be zero. The principal ideal (x) lies in C and has nonzero product with J, so minimality gives $C = (x)$. Also $J(Jx) = J^2x = Jx \neq 0$. The ideal Jx lies in (x) and again has nonzero product with J, so minimality now gives $Jx = (x)$. Hence $x = ax$ for some $a \in J$. Because $J \subseteq N$, the element a is nilpotent. Repeated substitution gives $x = a^r x = 0$ for large r, contradicting $Jx \neq 0$. Therefore $J = 0$, as claimed. □

This proof did not assume that A was Noetherian. Using Proposition 9.35 here would be circular, because the next theorem is what first proves that an Artinian ring is Noetherian.

Definition 9.44. The *Krull dimension* of a nonzero ring is the supremum of the integers d for which there exists a strict chain $P_0 \subsetneq P_1 \subsetneq \cdots \subsetneq P_d$ of primes. The chain has d inclusions. Thus dimension zero means that all primes are maximal. Lecture 12 studies longer chains systematically.

Theorem 9.45 (Artinian rings and dimension zero; AM 8.5)**.** *A nonzero ring is Artinian if and only if it is Noetherian and has dimension zero. Equivalently, allowing the zero ring, it is Artinian if and only if it is Noetherian and every prime ideal is maximal.*

Proof. Suppose A is Artinian. All primes are maximal, and the distinct maximal ideals form a finite list $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$. Their intersection is the nilradical N, and $N^k = 0$ for some k. Since $\mathfrak{m}_1 \cdots \mathfrak{m}_r \subseteq N$, we have

$$\mathfrak{m}_1^k \cdots \mathfrak{m}_r^k = (\mathfrak{m}_1 \cdots \mathfrak{m}_r)^k = 0.$$

This is a product of finitely many maximal ideals, with repetitions, equal to zero. Corollary 9.32 says the Artinian ring A is Noetherian.

Conversely, suppose A is Noetherian and every prime is maximal. Every prime is then minimal as well. Lemma 9.38 gives finitely many primes $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$. Their intersection is N, and $N^k = 0$ by Corollary 9.35. Thus the same displayed product is zero. Corollary 9.32, now applied starting from the Noetherian condition, proves that A is Artinian. The zero ring satisfies both chain conditions directly. □

Example 9.46 (Why the Noetherian assumption is necessary)**.** Let

$$B = k[x_1, x_2, \ldots]/(x_i x_j : \ i, j \geq 1).$$

Write $\bar{x}_i$ for the image of x_i. Every prime contains every $\bar{x}_i$, since $\bar{x}_i^2 = 0$. Thus $\mathfrak{m} = (\bar{x}_1, \bar{x}_2, \ldots)$ is the only prime; its quotient is k. However, $\mathfrak{m}$ cannot be finitely generated: it is an infinite-dimensional k-vector space with basis $\bar{x}_i$, and multiplication by elements of B on $\mathfrak{m}$ uses only their constant terms. Consequently B is zero-dimensional but not Noetherian, and hence not Artinian. Directly, the ideals spanned by $\bar{x}_n, \bar{x}_{n+1}, \ldots$ give a strictly descending chain.

Proposition 9.47 (Powers of the maximal ideal; AM 8.6)**.** *For a Noetherian local ring $(A, \mathfrak{m})$, exactly one of the following occurs: $\mathfrak{m}^n \neq \mathfrak{m}^{n+1}$ for every $n \geq 1$; or $\mathfrak{m}^n = 0$ for some $n \geq 1$, in which case A is Artinian.*

Proof. If two successive powers agree, the finitely generated module $M = \mathfrak{m}^n$ satisfies $\mathfrak{m}M = M$. Nakayama's lemma gives $M = 0$. Every prime then contains $\mathfrak{m}^n$, hence contains $\mathfrak{m}$ by taking radicals, and so equals $\mathfrak{m}$. Theorem 9.45 gives Artinianity. Conversely, a zero power makes all later powers zero, so strict descent at every stage is impossible. These alternatives are therefore both exhaustive and mutually exclusive. □

Theorem 9.48 (Structure of Artinian rings; AM 8.7)**.** *Every nonzero Artinian ring is a finite product of nonzero Artinian local rings. The factors are uniquely determined up to permutation and isomorphism. More precisely, its canonical local factors are $A_{\mathfrak{m}}$, one for each maximal ideal $\mathfrak{m}$. The zero ring may be regarded as the empty product.*

Proof. **Existence.** List the maximal ideals as $\mathfrak{m}_1, \ldots, \mathfrak{m}_r$. The preceding proof gives $\prod_i \mathfrak{m}_i^k = 0$ for some k. The ideals $\mathfrak{m}_i^k$ are pairwise comaximal. To check this explicitly, for $i \neq j$ choose $u \in \mathfrak{m}_i$ and $v \in \mathfrak{m}_j$ with $u + v = 1$. In the expansion of $(u+v)^{2k-1} = 1$, each monomial has either at least k copies of u or at least k copies of v. Thus $1 \in \mathfrak{m}_i^k + \mathfrak{m}_j^k$. For pairwise comaximal ideals, intersection equals product and the Chinese remainder map is an isomorphism. Therefore

$$A \longrightarrow \prod_{i=1}^{r} A/\mathfrak{m}_i^k, \qquad a \longmapsto (a + \mathfrak{m}_i^k)_i$$

is an isomorphism. Each factor is Artinian as a quotient. Its maximal ideals correspond to maximal ideals of A containing $\mathfrak{m}_i^k$. Such an ideal must contain $\mathfrak{m}_i$ and hence equal it. Thus each factor is local and nonzero.

The factors are canonical. For a product $B_1 \times \cdots \times B_r$ of nonzero local rings, let e_i be the element with 1 in position i and 0 elsewhere. Then $\sum_i e_i = 1$ and $e_i e_j = 0$ for $i \neq j$. In a prime quotient exactly one e_i survives: at least one must survive because their sum is 1, and at most one can survive because their pairwise products are zero. Thus primes of the product are exactly primes in a single factor, with all other factors included in the ideal. In particular, maximal ideals correspond one-to-one with the factors' unique maximal ideals.

Localize at the maximal ideal corresponding to B_i. The element e_i lies outside it, so becomes invertible. Since $e_i^2 = e_i$, an invertible e_i must equal 1, forcing every e_j with $j \neq i$

to vanish. This kills the other factors. In the surviving factor, the inverted elements are precisely those outside its maximal ideal, which are already units. Hence this localization is B_i. Any product decomposition into nonzero local rings therefore has exactly the factors $A_{\mathfrak{m}}$, independently of its construction. This proves uniqueness. □

Example 9.49 (Computing the local pieces)**.** The ring $\mathbb{Z}/60\mathbb{Z}$ has maximal ideals corresponding to $2,3,5$, and

$$\mathbb{Z}/60\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}.$$

One can see the factor projections inside the ring through the idempotents $45, 40, 36$: modulo $4, 3, 5$ they have coordinates $(1,0,0), (0,1,0), (0,0,1)$ respectively. Their sum is $121 \equiv 1 \pmod{60}$, and their pairwise products vanish modulo 60. The first factor retains a nilpotent element, the class of 2, while the other two factors are fields. Geometrically, a finite collection of points may carry extra nilpotent information at individual points.

Proposition 9.50 (Principal ideals in an Artinian local ring; AM 8.8)**.** *Let $(A, \mathfrak{m})$ be Artinian local, with residue field $k = A/\mathfrak{m}$. The following are equivalent: every ideal is principal; $\mathfrak{m}$ is principal;* $\dim_k(\mathfrak{m}/\mathfrak{m}^2) \leq 1$.

Proof. If every ideal is principal then $\mathfrak{m}$ is principal; if $\mathfrak{m} = (x)$ then its image spans $\mathfrak{m}/\mathfrak{m}^2$, whose dimension is at most one. Conversely, Artinianity implies Noetherianity, so $\mathfrak{m}$ is finitely generated and Nakayama's lemma applies. If the vector-space dimension is zero, $\mathfrak{m} = \mathfrak{m}^2$ and Nakayama gives $\mathfrak{m} = 0$. Then A is a field and its two ideals are principal. If the dimension is one, lift a basis vector to $x \in \mathfrak{m}$; Nakayama gives $\mathfrak{m} = (x)$.

The ideal $\mathfrak{m}$ is nilpotent. For a nonzero proper ideal I, choose the largest integer r with $I \subseteq \mathfrak{m}^r$; such a largest integer exists because sufficiently high powers are zero. Then some $y \in I$ lies outside $\mathfrak{m}^{r+1}$. Since $\mathfrak{m}^r = (x^r)$, write $y = ax^r$. If $a \in \mathfrak{m}$, then $y \in \mathfrak{m}^{r+1}$, a contradiction. Hence a is a unit and $x^r = a^{-1}y \in I$. We have $(x^r) \subseteq I \subseteq \mathfrak{m}^r = (x^r)$, so $I = (x^r)$. Finally, $0 = (0)$ and $A = (1)$ are principal as well. □

Example 9.51 (One infinitesimal direction or two)**.** In $k[t]/(t^4)$ every ideal is (t^r) for $0 \leq r \leq 4$, and the ring has length 4. In $k[x,y]/(x,y)^2$, the maximal ideal has square zero and k-basis x, y. Therefore $\dim_k \mathfrak{m}/\mathfrak{m}^2 = 2$, and it cannot be principal. Every line $k(ax+by)$ in that two-dimensional space is an ideal, because multiplication by any element of the ring acts on $\mathfrak{m}$ by its constant term. Even an Artinian ring can have infinitely many ideals when the residue field is infinite.

Length counts infinitesimal information, not just points

In $B = k[x,y]/(x^2, xy, y^3)$, every monomial reduces to a linear combination of

$$1, \quad x, \quad y, \quad y^2.$$

These monomials are independent because none belongs to the monomial ideal (x^2, xy, y^3). Thus $\dim_k B = 4$. The ideal $\mathfrak{m} = (x,y)$ satisfies $\mathfrak{m}^2 = (y^2)$ and $\mathfrak{m}^3 = 0$, and $B/\mathfrak{m} = k$. Every prime contains x, y, since they are nilpotent, so $\operatorname{Spec} B$ has exactly one point. Nevertheless

$$\operatorname{length}_B(B) = \dim_k(B/\mathfrak{m}) + \dim_k(\mathfrak{m}/\mathfrak{m}^2) + \dim_k(\mathfrak{m}^2) = 1 + 2 + 1 = 4.$$

This calculation will reappear as the local length of an intersection.

For a contrasting calculation with two points, take

$$C = k[t]/\big(t^2(t-1)^3\big).$$

The ideals (t^2) and $((t-1)^3)$ are comaximal, so

$$C \cong k[t]/(t^2) \times k[t]/((t-1)^3).$$

The two local factors have lengths two and three. Hence C has dimension and length five over k, distributed between two points. The nilradical removes the powers and gives $C/\sqrt{0} \cong k \times k$. Taking the reduced ring preserves the two underlying points and discards three units of length. This is why replacing an ideal by its radical is unsuitable when multiplicity is part of the question.

9.7 Finiteness of spectra and fibers

The exercises at the end of this lecture make precise what “finitely many points” means algebraically. Recall that $\operatorname{Spec} A$ is the set of prime ideals and that its Zariski closed subsets are

$$V(I) = \{P : I \subseteq P\}.$$

A space is *discrete* when every singleton is open, equivalently when every subset is open. Its singleton sets are then also closed, since their complements are open. These notions will be used again in Lectures 15–16.

For a map of sets $h : Y \to X$, its *fiber over* x is the inverse image $h^{-1}(\{x\})$. If Y is topological, the fiber has the *subspace topology*: its open subsets are its intersections with open subsets of Y. A ring map $f : A \to B$ is *of finite type* if B has finitely many algebra generators over the image of A; it is *finite* if B has finitely many A-module generators. The map is *integral* if every element of B satisfies a monic equation with coefficients in the image of A. These are properties of the algebra map, even when f is not injective. The induced map on spectra sends Q to $f^{-1}(Q)$.

The recurring distinction is between a finite set of points, finite vector-space dimension, and finite generation of ideals. They become equivalent only after the appropriate hypotheses are supplied. In the next lecture, valuation rings and integral extensions will show how these finiteness conditions control curves and arithmetic.

9.8 Exercises

Exercise 9.1 (Surjective and injective endomorphisms; AM, Chapter 6, Exercise 1)**.** Show that a surjective endomorphism of a Noetherian module is an isomorphism. Show that an injective endomorphism of an Artinian module is an isomorphism.

Exercise 9.2 (A restricted maximal condition suffices; AM, Chapter 6, Exercise 2)**.** Let M be an A-module. Suppose every nonempty collection of finitely generated submodules of M has a maximal element. Prove that M is Noetherian.

Exercise 9.3 (Two quotient tests; AM, Chapter 6, Exercise 3)**.** Let M be an A-module and $N_1, N_2 \subseteq M$ submodules. If M/N_1 and M/N_2 are Noetherian, prove that $M/(N_1 \cap N_2)$ is Noetherian. Prove the Artinian version as well.

Exercise 9.4 (The cusp is Noetherian, but its local ring is not a principal ideal ring)**.** Let $B = k[x,y]/(y^2 - x^3)$ and let $A = B_{(x,y)}$. Prove that both rings are Noetherian, and that the maximal ideal of A needs two generators.

Exercise 9.5 (Cohen's finiteness criterion; AM, Chapter 7, Exercise 1)**.** In a non-Noetherian ring, show that the set of ideals which are not finitely generated has maximal elements and that each such maximal element is prime. Deduce that a ring is Noetherian if every prime ideal is finitely generated.

Exercise 9.6 (Invariant rings of finite groups; AM, Chapter 7, Exercise 5)**.** Let A be Noetherian, let B be a finitely generated A-algebra, and let a finite group G act on B by A-algebra automorphisms. Prove that $B^G = \{b : \sigma(b) = b \text{ for every } \sigma \in G\}$ is a finitely generated A-algebra.

Exercise 9.7 (Fields finitely generated as rings; AM, Chapter 7, Exercise 6)**.** Prove that a field which is finitely generated as a ring is a finite field.

Exercise 9.8 (The annihilator ring; AM, Chapter 6, Exercise 4)**.** Let M be a Noetherian A-module and put $I = \mathrm{Ann}_A(M)$. Prove that A/I is Noetherian. Decide whether the analogous assertion for Artinian modules is true.

Exercise 9.9 (A discrete spectrum; AM, Chapter 8, Exercise 2)**.** For a Noetherian ring A, prove the equivalence of: A is Artinian; $\mathrm{Spec}\,A$ is finite and discrete; $\mathrm{Spec}\,A$ is discrete.

Exercise 9.10 (Finite type plus dimension zero; AM, Chapter 8, Exercise 3)**.** Let k be a field and let A be a finitely generated k-algebra. Prove that A is Artinian if and only if it is finite-dimensional as a k-vector space.

Exercise 9.11 (Discrete fibers need not make a map finite; AM, Chapter 8, Exercise 4)**.** Let $f : A \to B$ be a ring homomorphism of finite type. Prove that

(i) f finite implies every fiber of $\mathrm{Spec}\,B \to \mathrm{Spec}\,A$ is discrete;

(ii) the fibers are discrete if and only if, for every prime P of A, $B \otimes_A \kappa(P)$ is finite-dimensional over $\kappa(P)$;

(iii) either equivalent condition implies all fibers are finite sets.

Here $\kappa(P) = \mathrm{Frac}(A/P)$ is the residue field of A_P. Finally, if an integral ring map has finite fibers, must it be finite?

Exercise 9.12 (Ideals and radicals in a finite ring)**.** Let $A = \mathbb{Z}/12\mathbb{Z}$. Find its units, prime ideals, maximal ideals, nilradical, and Jacobson radical. Give a Chinese remainder decomposition and explain whether A is local, Noetherian, or Artinian.

Exercise 9.13 (Finite generation and Nakayama's lemma)**.** Show that $R = k[x_1, x_2, \ldots]$ is not Noetherian. Then give a local ring $(B, \mathfrak{m})$ and a nonzero module M satisfying $\mathfrak{m}M = M$, explaining why this does not contradict Nakayama's lemma.

Lecture 10
Valuations, integral dependence, and normalization

We study fractions by measuring their orders of vanishing and by asking whether they satisfy monic polynomial equations. The first construction leads to discrete valuation rings; the second leads to integral closure. The cusp $y^2 = x^3$ provides a concrete example in which integral closure adds a missing rational function. The remaining sections develop prime ideals in integral extensions, the trace of a field extension, and Dedekind domains. The main references are [1, Chapters 5 and 9]; the geometry of the cusp will return in Lecture 21.

10.1 Order of vanishing as arithmetic

Fix $a \in k$. Every nonzero polynomial has a unique expression

$$f(t) = (t-a)^r u(t), \qquad u(a) \neq 0.$$

For a nonzero rational function define

$$v_a(f/g) = r - s,$$

where r and s are the respective exponents of $t-a$ in numerator and denominator. This is independent of the representation: equality $f/g = f'/g'$ implies $fg' = f'g$, and the exponents on the two sides agree. If $v_a(f/g) = m > 0$, we say that the rational function has a *zero of order m at a*. If the value is $-m < 0$, it has a *pole of order m at a*. Value zero means that, after cancellation, the function has a nonzero value at a.

For example,

$$v_0\left(\frac{t^3(1+t)}{t^2(2+t)}\right) = 1$$

when $2 \neq 0$ in k. If $\operatorname{char} k = 2$, the denominator has an additional factor t and the value is 0. Hypotheses on coefficients matter even in a simple order calculation.

Definition 10.1. A *discrete valuation* on a field K is a surjective map $v : K^\times \to \mathbb{Z}$ such that

$$v(ab) = v(a) + v(b), \qquad v(a+b) \geq \min\{v(a), v(b)\} \quad (a+b \neq 0).$$

Set $v(0) = +\infty$, where $+\infty$ is larger than every integer and $n + (+\infty) = +\infty$. Its *valuation ring* is

$$V = \{a \in K : v(a) \geq 0\},$$

where 0 is included by convention. Such a ring is called a *discrete valuation ring*, or DVR.

The valuation identities give

$$v(1) = v(1) + v(1) = 0, \quad 2v(-1) = v(1) = 0,$$
$$v(-a) = v(a), \qquad\qquad v(z^{-1}) = -v(z).$$

Here the first equality forces $v(1) = 0$. Thus V is a subring, and

$$z \in K^\times \quad \Longrightarrow \quad z \in V \text{ or } z^{-1} \in V.$$

If $z \notin V$, the expression $z = 1/z^{-1}$ lies in $\mathrm{Frac}(V)$, so $K = \mathrm{Frac}(V)$. A domain with this either-z-or-z^{-1} property is called a *valuation ring*. It need not be discrete or Noetherian.

An element $\pi \in V$ with $v(\pi) = 1$ is called a *uniformizer*; one exists by surjectivity of v.

Proposition 10.2. *Choose a uniformizer π. Every nonzero $a \in V$ can be written $a = u\pi^r$ with u a unit and $r = v(a) \geq 0$. The unique maximal ideal is (π), and every nonzero ideal is (π^r) for a unique $r \geq 0$.*

Proof. Put $u = a\pi^{-r}$. Then $v(u) = v(u^{-1}) = 0$, so both u and its inverse lie in V. Thus the units are exactly the elements of value zero, and the nonunits are exactly those of value at least one, namely (π).

For a nonzero ideal I, the values of its nonzero elements are a nonempty subset of $\mathbb{N}$ and have a least member r. Choose $a \in I$ of value r. Every $b \in I$ has $v(b/a) \geq 0$, so $b \in (a)$. Hence $I = (a) = (\pi^r)$. Different powers have different least values, proving uniqueness. □

The proposition proves that a DVR is Noetherian and local. Its only primes are (0) and (π): for $r > 1$, the ideal (π^r) is not prime because $\pi \cdot \pi^{r-1}$ belongs to it while neither factor does. Thus its Krull dimension, defined in Lecture 9, is one.

The following examples identify these orders with the familiar localizations:

$$k[t]_{(t-a)} \quad \text{with uniformizer } t-a, \qquad \mathbb{Z}_{(p)} \quad \text{with uniformizer } p.$$

For the first example, a fraction of nonnegative order can have all factors $t-a$ canceled from its denominator; the remaining denominator is nonzero at a, as required in $k[t]_{(t-a)}$. Conversely every fraction in this localization has nonnegative order. In the second example, $v_p(m/n)$ is the exponent of p in m minus its exponent in n. Canceling common powers of p shows that its nonnegative-value ring is exactly $\mathbb{Z}_{(p)}$.

10.2 Monic equations and finite modules

Definition 10.3. For an inclusion $A \subseteq B$, an element $x \in B$ is *integral over* A if

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0 \qquad (a_i \in A)$$

for some positive integer n. The leading coefficient is required to be 1. The extension is integral if each element is integral. For a homomorphism $A \to B$ that need not be injective, integrality means integrality over the image of A.

For a reduced fraction r/s with $s > 0$, a monic equation over $\mathbb{Z}$ gives

$$\begin{aligned} & (r/s)^n + a_1 (r/s)^{n-1} + \cdots + a_n = 0 \\ \Longrightarrow \quad & r^n + a_1 r^{n-1} s + \cdots + a_n s^n = 0 \\ \Longrightarrow \quad & s \mid r^n, \quad \gcd(r,s) = 1 \Longrightarrow s = 1. \end{aligned}$$

Consequently a rational number is integral over $\mathbb{Z}$ exactly when it is an integer; for example, $1/2$ is algebraic over $\mathbb{Q}$ but not integral over $\mathbb{Z}$.

Recall that an R-module M is *faithful* if $\operatorname{Ann}_R(M) = 0$: the only element of R killing every member of M is zero. For example, a ring is faithful as a module over itself because a scalar killing 1 is zero. In the following criterion, faithfulness is required over $A[x]$, whereas finite generation is required over A.

Proposition 10.4 (The four finite-module tests; AM 5.1)**.** *For $x \in B$, the following conditions are equivalent:*

(i) *x is integral over A;*
(ii) *$A[x]$ is finitely generated as an A-module;*
(iii) *$A[x]$ lies in a subring $C \subseteq B$ that is finitely generated as an A-module;*
(iv) *there exists a faithful $A[x]$-module M that is finitely generated as an A-module.*

Proof. For (i)⇒(ii), the monic equation gives

$$x^n \in \sum_{j=0}^{n-1} A x^j \quad \Longrightarrow \quad A[x] = \sum_{j=0}^{n-1} A x^j,$$

by successive substitution in every higher power. For (ii)⇒(iii), take $C = A[x]$. For (iii)⇒(iv), take $M = C$: if $cC = 0$ for $c \in A[x]$, then $c = c \cdot 1 = 0$.

For (iv)⇒(i), write $M = Am_1 + \cdots + Am_r$ and

$$x m_i = \sum_j a_{ij} m_j, \qquad D = (a_{ij}), \qquad \mathbf{m} = (m_1, \ldots, m_r)^{\mathsf{t}}.$$

The adjugate identity yields

$$\begin{aligned} (xI - D)\mathbf{m} &= 0, \\ \operatorname{adj}(xI - D)(xI - D)\mathbf{m} &= 0, \\ \det(xI - D)\mathbf{m} &= 0. \end{aligned}$$

Thus the monic polynomial $P(T) = \det(TI - D) \in A[T]$ satisfies $P(x)M = 0$. Faithfulness gives $P(x) = 0$. If $M = 0$, faithfulness forces $A[x] = 0$, where integrality is immediate. □

Corollary 10.5 (Finitely many integral generators; AM 5.2)**.** *If $x_1, \dots, x_r \in B$ are integral over A, then $A[x_1, \dots, x_r]$ is a finite A-module.*

Proof. Choose a monic equation of degree d_i for each x_i. Reduce powers of each variable using its equation. The finitely many products

$$x_1^{e_1} \cdots x_r^{e_r}, \qquad 0 \le e_i < d_i,$$

span the whole algebra. They need not be linearly independent; finite generation is the assertion. □

Corollary 10.6 (Integral elements form a ring; AM 5.3)**.** *The integral elements of B over A form a subring C containing A.*

Proof. An element $a \in A$ solves $T - a = 0$. If x, y are integral, the ring $A[x, y]$ is finite over A by Proposition 10.5. It contains $x + y$, $x - y$ and xy. Apply condition (iii) of Proposition 10.4 to each of these elements. □

This subring C is the *integral closure of A in B*. We call A *integrally closed in B* if $C = A$. A domain is *integrally closed*, or *normal*, when it is integrally closed in its own fraction field.

Corollary 10.7 (Transitivity; AM 5.4)**.** *If $A \subseteq B \subseteq C$, B is integral over A, and C is integral over B, then C is integral over A.*

Proof. Let $x \in C$ satisfy

$$x^n + b_1 x^{n-1} + \cdots + b_n = 0, \qquad B_0 = A[b_1, \dots, b_n].$$

By Corollary 10.5, write $B_0 = \sum_{i=1}^{r} A u_i$. The equation for x then gives

$$B_0[x] = \sum_{j=0}^{n-1} B_0 x^j = \sum_{i=1}^{r} \sum_{j=0}^{n-1} A(u_i x^j).$$

This is a finite A-module containing x. Proposition 10.4 makes x integral over A. Only the finitely many coefficients b_i were needed; B itself need not be a finite A-module. □

Corollary 10.8 (Closing twice adds nothing; AM 5.5)**.** *If C is the integral closure of A in B, then C is integrally closed in B.*

Proof. An element of B integral over C is integral over A by transitivity and thus already lies in C. □

An algebra generated by finitely many elements is called *of finite type*; an algebra finite as a module is called *finite*. The results above prove

$$\text{finite} \iff \text{finite type and integral.}$$

The adjective "finite" is about a module-generating list, not about the number of elements in the ring.

Proposition 10.9 (Quotients and localization; AM 5.6)**.** *Suppose $A \subseteq B$ is integral.*

(i) If $J \subseteq B$ is an ideal and $I = J \cap A$, then $A/I \subseteq B/J$ is integral.

(ii) If $S \subseteq A$ is multiplicatively closed, then $S^{-1}A \subseteq S^{-1}B$ is integral.

Proof. For (i), reduce a monic equation modulo J; the coefficient images lie in A/I. For (ii), a monic equation $b^n + a_1 b^{n-1} + \cdots + a_n = 0$ gives

$$(b/s)^n + (a_1/s)(b/s)^{n-1} + \cdots + a_n/s^n = 0.$$

All coefficients lie in $S^{-1}A$. The localized map remains injective: if a/s maps to zero, then some $t \in S$ satisfies $ta = 0$ in B, hence also in A. □

Integral, finite, and finite type are different conditions

For an A-algebra B, "finite type" means finitely generated as an algebra, while "finite" means finitely generated as an A-module. If $B = A[b_1, \ldots, b_r]$ and each b_i satisfies a monic equation of degree d_i, then the finitely many products

$$b_1^{e_1} \cdots b_r^{e_r}, \qquad 0 \le e_i < d_i,$$

span B as an A-module. Reduce a power $b_i^{d_i}$ by its monic equation and repeat. Thus integral plus finite type implies finite. Conversely, a finite algebra is integral by the determinant argument applied to multiplication by an element.

The polynomial algebra $A[t]$ is finite type, but when $A \neq 0$ it is not finite as an A-module: any finite list of polynomials has a degree bound which its A-linear span cannot exceed. An integral algebra need not be finite type. For example, the algebraic closure $\overline{\mathbf{F}_p}$ is integral over $\mathbf{F}_p$ because each of its elements is algebraic. It is not a finite-dimensional vector space: it contains finite subfields of arbitrarily large degrees. A finite-dimensional extension of $\mathbf{F}_p$ would have only finitely many elements. Equivalently one may take the union of the fields $\mathbf{F}_{p^{n!}}$ inside an algebraic closure.

One further useful fact is that an integral extension preserves the property of being a field when the larger ring is a domain. If $A \subseteq B$ is integral and B is a field, for $0 \neq a \in A$ the element $a^{-1} \in B$ satisfies

$$a^{-n} + c_1 a^{-(n-1)} + \cdots + c_n = 0.$$

Multiplying by a^{n-1} expresses a^{-1} as an element of A. Conversely, if A is a field and $0 \neq b \in B$, choose a monic relation of smallest degree for b. Its constant coefficient is nonzero, since otherwise cancellation of b in the domain would give a smaller relation. Solving the equation for b^{-1} places the inverse in $A[b] \subseteq B$. These proofs explain the field criterion used in lying over.

10.3 Polynomial factorization and integral closure

The normalization calculation below needs a fact about polynomial fractions: a reduced denominator cannot survive a monic equation. We establish the factorization results that make this precise. They also apply to polynomial rings in several variables, which will occur in the dimension and geometry lectures.

Recall that in a domain two nonzero elements are *associates* if one is a unit times the other. A nonzero nonunit p is *irreducible* if $p = ab$ forces a or b to be a unit; it is *prime* if $p \mid ab$ forces $p \mid a$ or $p \mid b$. A *unique factorization domain*, abbreviated UFD, is a domain in which every nonzero nonunit is a finite product of irreducibles, uniquely up to their order and replacement by associates. In a UFD, irreducibles are prime: compare factorizations of the two sides of $ab = pc$.

Proposition 10.10 (Factorization in one variable)**.** *For every field k, the ring $k[t]$ is a UFD.*

Proof. Lecture 1 proved, using division with remainder, that every ideal of $k[t]$ is principal. Let p be irreducible and suppose $p \nmid f$. Write $(p, f) = (d)$. Since d divides p, irreducibility makes d either a unit or an associate of p. The latter would imply $p \mid f$, so d is a unit. Consequently there are $u, v \in k[t]$ with $up + vf = 1$. If $p \mid fg$, multiplying this identity by g shows $p \mid g$. Thus every irreducible is prime.

Existence of factorization follows by induction on degree: a nonconstant polynomial is irreducible or splits into two factors of smaller positive degrees. For uniqueness, an irreducible in one factorization divides the product in the other, hence divides one factor there. Those irreducible factors are associates. Cancel them in the domain and repeat. Units are the nonzero constants. □

Theorem 10.11 (Gauss's lemma and polynomial factorization)**.** *If R is a UFD, then $R[t]$ is a UFD. In particular, $k[x_1, \ldots, x_n]$ is a UFD for every field k and every finite n.*

Proof. Let $K = \mathrm{Frac}(R)$. A *greatest common divisor* of finitely many nonzero elements is an element dividing all of them and divisible by every common divisor. In a domain it is unique up to multiplication by a unit when it exists. For a nonzero polynomial $f \in R[t]$, its *content* $c(f)$ is a greatest common divisor of its nonzero coefficients, determined up to a unit. It exists by taking the minimum exponent of each irreducible in their factorizations. Write $f = c(f) f_0$. The polynomial f_0 is *primitive in the content sense*: no irreducible of R divides every coefficient. In this proof, "primitive" always has this meaning. It says that the coefficient gcd is a unit. The convention in Exercise 2.4, that the coefficients generate the

unit ideal, is stronger over a general UFD. For instance $uT + v$ over $k[u,v]$ has coefficient gcd 1, but its coefficient ideal (u,v) is proper. The two conventions agree over a principal ideal domain.

Step 1: products of primitive polynomials are primitive. If an irreducible $p \in R$ divided every coefficient of $f_0 g_0$, reduce modulo (p). Since p is prime, $R/(p)$ is a domain. Each factor gives a nonzero polynomial there, whereas their product would be zero. This is impossible: the product of the leading coefficients of two nonzero polynomials over a domain is nonzero. Thus $f_0 g_0$ is primitive. It follows that $c(fg)$ is associated to $c(f)c(g)$.

Step 2: scalar multiples of primitive polynomials are rigid. Suppose primitive polynomials $f, h \in R[t]$ satisfy $f = \gamma h$ for $\gamma \in K^\times$. Write $\gamma = a/b$ with numerator and denominator having no common irreducible factor. From $bf = ah$, every irreducible factor of b, with its full exponent, must divide every coefficient of h. Primitivity forces b to be a unit. Then $f = (a/b)h$ and primitivity of f forces a to be a unit as well. Hence $\gamma \in R^\times$.

Step 3: transfer factorization from the fraction field. Every nonzero $H \in K[t]$ can be written $H = \alpha H_0$ with $\alpha \in K^\times$ and $H_0 \in R[t]$ primitive: clear the finitely many coefficient denominators and then divide out the content. Factor a primitive $f \in R[t]$ in the UFD $K[t]$, and perform this replacement on each positive-degree irreducible factor. We obtain

$$f = \gamma H_1 \cdots H_r,$$

where each H_i is primitive and irreducible over K. Step 1 makes the product primitive, so Step 2 makes γ a unit of R.

Each H_i is irreducible in $R[t]$: a factorization into two positive-degree factors would remain a factorization over K, and a constant nonunit factor would contradict primitivity. Conversely, a positive-degree irreducible of $R[t]$ is primitive, since otherwise its content gives a nontrivial factorization. It must be irreducible over K, since the preceding transfer procedure would otherwise factor it in $R[t]$.

Finally factor $c(f)$ in R and the primitive part as above. This proves existence in $R[t]$. For uniqueness, the constant irreducible factors are determined by the content and uniqueness in R. The positive-degree primitive factors are determined up to K-scalar multiples by uniqueness in $K[t]$; Step 2 makes these scalars units of R. This proves uniqueness in $R[t]$. Starting with the field k and adjoining one variable at a time proves the final assertion. □

Proposition 10.12 (Unique factorization implies integral closure)**.** *Every UFD is integrally closed in its fraction field. In particular, every polynomial ring in finitely many variables over a field is integrally closed.*

Proof. For $a/b \in \mathrm{Frac}(R)$, cancel common irreducible factors and take $b \neq 0$. A monic equation of degree m gives

$$a^m + c_1 a^{m-1} b + \cdots + c_m b^m = 0, \qquad c_i \in R,$$

so $b \mid a^m$. If an irreducible p divided b, primality would give

$$p \mid b \mid a^m \quad \Longrightarrow \quad p \mid a,$$

contrary to cancellation. Hence $b \in R^\times$ and $a/b \in R$. The fraction 0 already lies in R. Apply Theorem 10.11 for the polynomial-ring assertion. □

10.4 The integral closure of the cusp ring

Let $A = k[t^2, t^3] \subseteq k[t]$. Substitution $x \mapsto t^2$, $y \mapsto t^3$ gives

$$A \cong k[x,y]/(y^2 - x^3).$$

Here is a check of the kernel. Divide a polynomial by the monic polynomial $y^2 - x^3$ in the variable y. Its remainder is $a(x) + yb(x)$, whose image is $a(t^2) + t^3 b(t^2)$. The first summand has even powers of t; the second has odd powers at least 3. These powers cannot cancel. Thus the remainder vanishes only if both polynomials are zero.

The semigroup of exponents generated by $2, 3$ is $\{0, 2, 3, 4, \ldots\}$. Consequently

$$t = \frac{t^3}{t^2} \in \mathrm{Frac}(A), \qquad t \notin A, \qquad t^2 \in A.$$

Thus $\mathrm{Frac}(A) = k(t)$, and t is integral over A, satisfying $T^2 - t^2 = 0$. The ring A is not normal.

Let $\overline{A}$ be its integral closure in $k(t)$. Since

$$k[t] = A + At,$$

Proposition 10.4 gives $k[t] \subseteq \overline{A}$. Conversely,

$$z \in \overline{A} \implies z \text{ integral over } k[t] \implies z \in k[t],$$

where the last implication uses Proposition 10.12. Therefore $\overline{A} = k[t]$.

For a domain R, replacing R by its integral closure in $\mathrm{Frac}(R)$ is called *normalization*. Thus the normalization of A is $k[t]$. Geometrically, the inclusion sends a parameter value $a \in k$ to the point (a^2, a^3) on the cusp. We have added the missing integral rational function t while keeping the fraction field unchanged. Lecture 21 studies this construction for curves in more detail.

For an inclusion $A \subseteq B$, its *conductor* is the set $\mathfrak{c} = \{a \in A : aB \subseteq A\}$. It is an ideal of both rings: closure under A-multiplication is immediate, and if $a \in \mathfrak{c}$ and $b \in B$, then $ab \in A$ and $(ab)B \subseteq aB \subseteq A$.

10.5 Lying over and going up

These theorems describe what happens to prime ideals when we enlarge a ring integrally. The relation $\mathfrak{q} \cap A = \mathfrak{p}$ is read as "$\mathfrak{q}$ lies over $\mathfrak{p}$." The two ideals belong to different rings; contraction is what makes them comparable.

Proposition 10.13 (Fields in an integral extension; AM 5.7)**.** *If $A \subseteq B$ are domains and B is integral over A, then A is a field if and only if B is a field.*

Proof. Suppose A is a field and $0 \neq b \in B$. Choose a monic equation for b of smallest degree. Its constant term a_n is nonzero: otherwise cancel b in the domain to obtain a shorter equation. Rearrangement gives

$$b^{-1} = -a_n^{-1}(b^{n-1} + a_1 b^{n-2} + \cdots + a_{n-1}) \in B.$$

Conversely, if B is a field, $0 \neq a \in A$ has inverse in B. A monic equation for a^{-1}, multiplied by a^{n-1}, expresses a^{-1} as an A-linear combination of $1, a, \ldots, a^{n-1}$. Hence $a^{-1} \in A$. □

Corollary 10.14 (Maximal ideals correspond under contraction; AM 5.8)**.** *For an integral extension $A \subseteq B$ and a prime $\mathfrak{q} \subseteq B$, the ideal $\mathfrak{q}$ is maximal if and only if $\mathfrak{p} = \mathfrak{q} \cap A$ is maximal.*

Proof. The quotient inclusion $A/\mathfrak{p} \subseteq B/\mathfrak{q}$ is an integral extension of domains. Apply Proposition 10.13 and the criterion that an ideal is maximal exactly when its quotient is a field. □

Corollary 10.15 (Incomparability; AM 5.9)**.** *If $\mathfrak{q} \subseteq \mathfrak{q}'$ are primes of an integral extension B of A and $\mathfrak{q} \cap A = \mathfrak{q}' \cap A$, then $\mathfrak{q} = \mathfrak{q}'$.*

Proof. Let their common contraction be $\mathfrak{p}$ and localize by $S = A \setminus \mathfrak{p}$. Both primes are disjoint from S, so their extensions are primes of $S^{-1}B$ with contraction $\mathfrak{p}A_{\mathfrak{p}}$. Proposition 10.9 and Corollary 10.14 say these extended primes are maximal. One contains the other, so they are equal. Contracting to B recovers the original primes and proves equality. □

Theorem 10.16 (Lying over; AM 5.10)**.** *Every prime $\mathfrak{p}$ of A is the contraction of a prime of B when $A \subseteq B$ is integral.*

Proof. Set $S = A \setminus \mathfrak{p}$. The nonzero ring $A_{\mathfrak{p}}$ injects into $S^{-1}B$, so the latter has a maximal ideal $\mathfrak{n}$. Its contraction is maximal in $A_{\mathfrak{p}}$ by Corollary 10.14; therefore it is the unique maximal ideal $\mathfrak{p}A_{\mathfrak{p}}$. The contraction $\mathfrak{q}$ of $\mathfrak{n}$ to B is prime and has $\mathfrak{q} \cap A = \mathfrak{p}$. □

Theorem 10.17 (Going up; AM 5.11)**.** *Let $A \subseteq B$ be integral. Suppose*

$$\mathfrak{p}_1 \subseteq \cdots \subseteq \mathfrak{p}_n \quad \text{and} \quad \mathfrak{q}_1 \subseteq \cdots \subseteq \mathfrak{q}_m \quad (m < n)$$

are prime chains in A and B, respectively, with $\mathfrak{q}_i \cap A = \mathfrak{p}_i$ for $i \leq m$. The second chain extends to primes $\mathfrak{q}_{m+1}, \ldots, \mathfrak{q}_n$ satisfying $\mathfrak{q}_i \cap A = \mathfrak{p}_i$ for all i.

Proof. It suffices to add one prime. The quotient inclusion $A/\mathfrak{p}_m \subseteq B/\mathfrak{q}_m$ is integral. By lying over, $\mathfrak{p}_{m+1}/\mathfrak{p}_m$ is the contraction of a prime in $B/\mathfrak{q}_m$. Its inverse image $\mathfrak{q}_{m+1}$ in B contains $\mathfrak{q}_m$ and contracts to $\mathfrak{p}_{m+1}$. Repeat this argument until the chain has length n. □

Example 10.18 (Primes can split)**.** In $\mathbb{Z} \subseteq \mathbb{Z}[i]$, the prime (5) has two distinct primes above it:

$$\mathfrak{q}_+ = (2+i), \qquad \mathfrak{q}_- = (2-i).$$

The quotients are $\mathbf{F}_5$, with i mapping to -2 and 2, respectively. Their product is (5). Incomparability says that neither prime can contain the other. Lying over gives existence, not uniqueness.

10.6 Localization, normality, and going down

Proposition 10.19 (Integral closure commutes with localization; AM 5.12)**.** *Let C be the integral closure of A in B, where $A \subseteq B$, and let $S \subseteq A$ be multiplicatively closed. Then $S^{-1}C$ is the integral closure of $S^{-1}A$ in $S^{-1}B$.*

Proof. Every element of $S^{-1}C$ is integral over $S^{-1}A$ by Proposition 10.9. For the converse let $z = b/s$ satisfy

$$z^n + \sum_{i=1}^{n} (a_i/s_i) z^{n-i} = 0, \qquad a_i \in A, \quad s, s_i \in S.$$

Set $t = s_1 \cdots s_n$ and $y = tb$. Multiplication by $(st)^n$ produces

$$y^n + c_1 y^{n-1} + \cdots + c_n = 0 \quad \text{in } S^{-1}B, \qquad c_i = a_i s^i t^i / s_i \in A.$$

This equality might not yet hold in B, because B may have S-torsion. Choose $u \in S$ annihilating its left side in B. Multiply the equality by u^n: the element uy then satisfies

$$(uy)^n + u c_1 (uy)^{n-1} + \cdots + u^n c_n = 0 \quad \text{in } B.$$

Thus $uy = utb \in C$ and $b/s = (utb)/(uts) \in S^{-1}C$. The additional factor u is what makes the proof valid for rings with zero divisors as well as domains. □

Proposition 10.20 (Normality is local; AM 5.13)**.** *For a domain A, the following are equivalent: A is integrally closed; every $A_{\mathfrak{p}}$ is integrally closed; every $A_{\mathfrak{m}}$ is integrally closed, where $\mathfrak{p}$ ranges over primes and $\mathfrak{m}$ over maximal ideals.*

Proof. The first condition implies the second by Proposition 10.19, and the second implies the third. Assume the third, and let $x \in \mathrm{Frac}(A)$ be integral over A. Its same equation is a monic equation over each $A_{\mathfrak{m}}$, so $x \in A_{\mathfrak{m}}$ for all $\mathfrak{m}$.

To see that $x \in A$, consider $I = \{a \in A : ax \in A\}$. If I were proper, choose a maximal ideal $\mathfrak{m} \supseteq I$. Membership $x \in A_{\mathfrak{m}}$ means $x = b/s$ with $s \notin \mathfrak{m}$. Then $sx = b \in A$, giving $s \in I \subseteq \mathfrak{m}$, a contradiction. Therefore $I = A$ and $x \in A$. □

Definition 10.21 (Integrality over an ideal: the convention in Chapter 5)**.** For an ideal $I \subseteq A$, say $x \in B$ is *integral over I in the Chapter 5 sense* if it solves a monic equation all of whose

nonleading coefficients belong to I. This is the convention used in AM 5.14–5.15. It differs from the later common convention requiring the coefficient of T^{n-i} to lie in I^i.

Lemma 10.22 (The radical description; AM 5.14)**.** *If C is the integral closure of A in B, the elements of B integral over I in this sense are exactly $\sqrt{IC}$, considered as a subset of $C \subseteq B$. In particular the set is closed under addition, subtraction, and multiplication.*

Proof. If $x^n + a_1 x^{n-1} + \cdots + a_n = 0$ with $a_i \in I$, then $x \in C$ and $x^n \in IC$, so $x \in \sqrt{IC}$.

Conversely, write $x^n = \sum_{j=1}^{r} a_j c_j$ with $a_j \in I$ and $c_j \in C$. Set $M = A[c_1, \ldots, c_r]$, a finite A-module containing 1. Multiplication by x^n maps M into IM, because $x^n \in IM \subseteq M$. Choose an A-generating list of size N for M and write the multiplication matrix with entries in I. The determinant argument from Proposition 10.4 gives

$$(x^n)^N + d_1 (x^n)^{N-1} + \cdots + d_N = 0, \qquad d_i \in I^i \subseteq I.$$

The equality follows by applying the annihilating determinant to $1 \in M$. Insert zero coefficients between the displayed powers to obtain a monic equation for x with all nonleading coefficients in I. □

The minimal polynomial was constructed in Proposition 9.20. We will also use its roots in a larger field. A polynomial *splits* over a field if it is a product of linear factors there; one can obtain such a field by adjoining roots successively. A root α has *multiplicity* e if $(T-\alpha)^e$ divides the polynomial but $(T-\alpha)^{e+1}$ does not. Thus a monic polynomial of degree d that splits has an expression

$$\prod_{j=1}^{d} (T - x_j),$$

where roots are repeated according to their multiplicities. Expanding the product shows that the coefficient of T^{d-i} is $(-1)^i$ times the sum of the products $x_{j_1} \cdots x_{j_i}$ over all i-element subsets of $\{1, \ldots, d\}$. This elementary coefficient formula is what the next proof uses.

Proposition 10.23 (Minimal-polynomial coefficients; AM 5.15)**.** *Let $A \subseteq B$ be domains, let A be integrally closed, and let $x \in B$ be integral over an ideal $I \subseteq A$ in the preceding sense. Then x is algebraic over $K = \operatorname{Frac}(A)$, and the nonleading coefficients of its monic minimal polynomial over K lie in $\sqrt{I}$.*

Proof. A monic equation over A already proves algebraicity over K. Work in a field containing $\operatorname{Frac}(B)$ and all roots of the minimal polynomial. List these roots with multiplicities as $x_1, \ldots, x_d$. Each is a root of the original monic equation: the minimal polynomial divides that equation in $K[T]$. Thus every x_j is integral over I.

Let C be the integral closure of A in this larger field. Lemma 10.22 puts every x_j in $\sqrt{IC}$. The coefficients of the minimal polynomial are signed sums of products of the x_j, hence also lie in $\sqrt{IC}$. In particular they are integral over A. They already belong to K, and normality gives that they belong to A. Finally an element $a \in A$ integral over I has a monic equation with coefficients in I, so $a^N \in I$ for some N. Therefore each coefficient belongs to $\sqrt{I}$. Repeated roots cause no problem: the symmetric-polynomial argument counts multiplicities. □

Theorem 10.24 (Going down; AM 5.16)**.** *Let $A \subseteq B$ be domains, A integrally closed and B integral over A. Suppose*

$$\mathfrak{p}_1 \supseteq \cdots \supseteq \mathfrak{p}_n, \qquad \mathfrak{q}_1 \supseteq \cdots \supseteq \mathfrak{q}_m \quad (m < n)$$

are prime chains in A and B, with $\mathfrak{q}_i \cap A = \mathfrak{p}_i$ for $i \le m$. The second chain extends downwards to primes $\mathfrak{q}_{m+1}, \dots, \mathfrak{q}_n$ with the corresponding contractions.

Proof. We prove one step, writing $\mathfrak{p}' \subseteq \mathfrak{p}$ and fixing $\mathfrak{q}$ over $\mathfrak{p}$. Let $D = B_{\mathfrak{q}}$. The crucial claim is

$$\mathfrak{p}' D \cap A = \mathfrak{p}'. \tag{*}$$

The inclusion from right to left is immediate. For the other inclusion take $0 \neq a \in \mathfrak{p}' D \cap A$. We can write $a = y/s$ with $y \in \mathfrak{p}' B$ and $s \in B \setminus \mathfrak{q}$ by collecting denominators in a finite sum. By Lemma 10.22, y is integral over $\mathfrak{p}'$. Proposition 10.23 says its minimal polynomial over $K = \operatorname{Frac}(A)$ has the form

$$T^r + u_1 T^{r-1} + \cdots + u_r, \qquad u_i \in \mathfrak{p}'.$$

Since $s = y/a$ and $a \in K^\times$, scaling the variable shows that the minimal polynomial of s is

$$T^r + v_1 T^{r-1} + \cdots + v_r, \qquad v_i = u_i / a^i.$$

This polynomial remains irreducible because scaling is an invertible substitution in $K[T]$. As s is integral over A, Proposition 10.23 with $I = A$ gives $v_i \in A$.

If $a \notin \mathfrak{p}'$, the equations $a^i v_i = u_i \in \mathfrak{p}'$ and primality imply $v_i \in \mathfrak{p}'$ for all i. Evaluating the polynomial at s yields $s^r \in \mathfrak{p}' B \subseteq \mathfrak{q}$, hence $s \in \mathfrak{q}$, a contradiction. Thus $a \in \mathfrak{p}'$, proving (*).

Now localize D by the images of $A \setminus \mathfrak{p}'$. Claim (*) ensures that the extension of $\mathfrak{p}' D$ is proper: if it contained 1, some element of $A \setminus \mathfrak{p}'$ would already lie in $\mathfrak{p}' D$. Choose a prime containing that proper ideal and contract it to D. Its contraction to A contains $\mathfrak{p}'$ and avoids $A \setminus \mathfrak{p}'$, hence equals $\mathfrak{p}'$. Finally primes of $B_{\mathfrak{q}}$ contract to primes of B contained in $\mathfrak{q}$. This gives the required $\mathfrak{q}' \subseteq \mathfrak{q}$. Repeat for each remaining step of the chain. □

The normality hypothesis entered at a precise point: it put the coefficients v_i back into A. Merely knowing that they lie in $\operatorname{Frac}(A)$ would not allow the primality argument.

10.7 Finite separable extensions and the trace

Definition 10.25. An algebraic element α over K is *separable* if its minimal polynomial has distinct roots in an algebraic closure of K. An algebraic extension L/K is separable if every element of L is separable over K. A *K-embedding* $\sigma : L \to \Omega$ is an injective field homomorphism that fixes each element of K. Its values on an algebraic element must be roots of the same minimal polynomial, since applying σ preserves every polynomial equation over K.

For a polynomial $f(T)=\sum_i a_i T^i$, its *formal derivative* is $f'(T)=\sum_{i\geq 1} i a_i T^{i-1}$, with the integer i acting as a scalar in the field. This definition satisfies the product rule, as follows by multiplying monomials. If $f=(T-\alpha)g$, then $f'(\alpha)=g(\alpha)$. Consequently a root is repeated precisely when it is a common root of f and f'. Equivalently, f and f' have a common nonconstant factor: their greatest common divisor is unchanged upon passage to a algebraic closure, since a Bezout identity for a coprime pair remains an identity after extending the field.

In characteristic zero a nonconstant irreducible f has nonzero derivative of smaller degree, so f and f' are coprime. All algebraic extensions in characteristic zero are therefore separable. The condition can fail in characteristic p. For example, let $L=\mathbf{F}_p(t)$ and $K=\mathbf{F}_p(t^p)$. The elements $1,t,\ldots,t^{p-1}$ are independent over K: after clearing denominators, a dependence would be a polynomial identity in t whose distinct summands have exponents in distinct residue classes modulo p. They also span, because $t^p\in K$ and the algebra $K[t]$ is a field by Proposition 9.20. Hence $[L:K]=p$, the minimal polynomial of t is T^p-t^p, and it is $(T-t)^p$ over L. This extension is not separable.

Definition 10.26. Let L/K be a finite field extension. For $z\in L$, multiplication $m_z: w\mapsto zw$ is a K-linear map $L\to L$. Its *field trace* is

$$\mathrm{Tr}_{L/K}(z)=\mathrm{tr}(m_z),$$

where the trace of a matrix is the sum of its diagonal entries. This is independent of the basis: the matrices in two bases are conjugate, and $\mathrm{tr}(UV)=\mathrm{tr}(VU)$, as one sees by summing their diagonal entries. In particular $\mathrm{tr}(P^{-1}MP)=\mathrm{tr}(M)$.

Example 10.27. For $L=\mathbb{Q}(\sqrt{2})$ and $z=a+b\sqrt{2}$, the matrix of m_z in the basis $1,\sqrt{2}$ is

$$\begin{pmatrix} a & 2b \\ b & a \end{pmatrix}.$$

Its trace is $2a$. There are two $\mathbb{Q}$-embeddings into $\mathbb{C}$, taking $\sqrt{2}$ to $\sqrt{2}$ and $-\sqrt{2}$, respectively. The sum of their values on z is again $2a$.

Proposition 10.28 (Trace and a separable extension)**.** *If L/K is finite separable of degree n and Ω is an algebraic closure of K, there are exactly n distinct K-embeddings $\sigma_i: L\to\Omega$, and*

$$\mathrm{Tr}_{L/K}(z)=\sum_{i=1}^{n}\sigma_i(z).$$

The map $\beta: L\times L\to K$, $\beta(x,y)=\mathrm{Tr}_{L/K}(xy)$, is a nondegenerate bilinear pairing. Here bilinear *means linear in each variable separately;* nondegenerate *means that for each nonzero x there is a y with $\beta(x,y)\neq 0$, and conversely with the variables interchanged.*

Proof. Choose finitely many generators and adjoin them successively. For one generator with minimal polynomial f, an embedding is determined by the root chosen as its image. Separability gives $\deg f$ distinct choices. At the next stage the minimal polynomial over the intermediate field divides the polynomial over K, so it also has distinct roots. Each existing embedding extends in as many ways as that next degree. The tower formula proves that there are exactly n embeddings.

For the trace calculation, *extension of scalars* from K to Ω replaces a K-vector space W by $\Omega \otimes_K W$. A basis of W gives the basis $1 \otimes w_i$ of this new Ω-vector space, and a linear map has the same matrix, now with entries regarded in Ω. Its trace is unchanged. In the one-generator case,

$$\Omega \otimes_K K[T]/(f) \cong \Omega[T]/(f) \cong \prod_{f(a)=0} \Omega.$$

The first isomorphism follows by sending $c \otimes h(T)$ to $ch(T)$ and imposing the relation $f(T) = 0$. The second is the Chinese remainder theorem, since the distinct linear factors generate pairwise comaximal ideals. Iterating along the tower gives

$$\Omega \otimes_K L \cong \prod_{i=1}^{n} \Omega, \qquad c \otimes z \longmapsto (c\sigma_i(z))_i.$$

Multiplication by z becomes the diagonal map with entries $\sigma_i(z)$. Adding those entries proves the trace formula.

Distinct embeddings are linearly independent as functions with values in Ω. Suppose, to the contrary, that $\sum_i c_i \sigma_i(z) = 0$ for every z, with a nonzero relation having the fewest nonzero coefficients. A one-term relation is impossible by evaluation at 1. Choose two terms, say $1, j$, and $a \in L$ with $\sigma_1(a) \neq \sigma_j(a)$. Subtract $\sigma_1(a)$ times the relation for z from the relation for az. The coefficient of σ_1 disappears and that of σ_j remains nonzero, contradicting minimality. In particular the sum of all embeddings, with every coefficient 1, is not the zero function, even in positive characteristic.

Choose z with nonzero trace. For any nonzero x, set $y = z/x$. Then $\beta(x, y) \neq 0$. Commutativity gives the same assertion with the variables reversed. Bilinearity follows from linearity of the matrix trace and the equalities $m_{x+x'} = m_x + m_{x'}$ and $m_{cx} = cm_x$ for $c \in K$. □

For a nondegenerate pairing on an n-dimensional vector space, every basis $u_1, \ldots, u_n$ has a *dual basis for the pairing*: vectors v_j with $\beta(u_i, v_j) = \delta_{ij}$, where δ_{ij} is 1 if $i = j$ and 0 otherwise. Indeed, the map $v \mapsto (\beta(u_i, v))_i$ from L to K^n is injective by nondegeneracy and hence bijective by equality of dimensions. Take the preimages of the standard coordinate vectors. In the quadratic example the basis dual to $1, \sqrt{2}$ is $1/2, \sqrt{2}/4$.

If A is a domain with fraction field K, an A-submodule of L of the form $Av_1 + \cdots + Av_n$, for a K-basis v_i of L, will be called a *free A-lattice*. The word here refers to a free module spanning L over K, with no assumption about a topology.

Proposition 10.29 (Bounding an integral closure by a lattice; AM 5.17)**.** *Let A be an integrally closed domain, $K = \mathrm{Frac}(A)$, L/K a finite separable extension, and B the integral closure of A in L. There exists a K-basis $v_1, \ldots, v_n$ of L such that*

$$B \subseteq Av_1 + \cdots + Av_n.$$

No Noetherian hypothesis is required for this containment.

Proof. Start with any K-basis $w_1, \ldots, w_n$. Clear the denominators in an algebraic equation for w_i to write

$$a_0 w_i^d + a_1 w_i^{d-1} + \cdots + a_d = 0, \qquad a_j \in A, \quad a_0 \neq 0.$$

Multiplying by a_0^{d-1} gives a monic equation for $a_0 w_i$. Thus we may rescale each basis vector and obtain a basis $u_1, \ldots, u_n$ with every $u_i \in B$.

Nondegeneracy of the trace pairing gives a dual basis $v_1, \ldots, v_n$ satisfying $\mathrm{Tr}(u_i v_j) = \delta_{ij}$. For $x \in B$, write $x = \sum_j c_j v_j$ with $c_j \in K$. Pairing with u_i gives

$$c_i = \mathrm{Tr}(x u_i).$$

Each image $\sigma_j(x u_i)$ of the integral element $x u_i$ satisfies the same monic equation over A, because σ_j fixes K and hence A. The trace, being their sum, is integral over A by Corollary 10.6. It lies in K, so normality of A gives $c_i \in A$. This proves the containment. □

If A is Noetherian, the containing free module is Noetherian and its submodule B is finite. This is one source of finiteness of normalization. Without the Noetherian assumption, containment in a finite free module alone does not establish finite generation.

Computing an integral closure in a quadratic field

Take $K = \mathbb{Q}(\sqrt{5})$. The element

$$\omega = (1 + \sqrt{5})/2$$

satisfies $\omega^2 - \omega - 1 = 0$, so it is integral over $\mathbb{Z}$, although it does not belong to $\mathbb{Z}[\sqrt{5}]$. In fact the full ring of integers is $\mathbb{Z}[\omega]$.

To prove the reverse inclusion, write an algebraic integer as $\alpha = a + b\sqrt{5}$ with $a, b \in \mathbb{Q}$. Its conjugate $a - b\sqrt{5}$ is also integral. Their sum and product are rational algebraic integers, hence integers, because $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$. Therefore

$$m = 2a \in \mathbb{Z}, \qquad n_0 = a^2 - 5b^2 \in \mathbb{Z}, \qquad 5(2b)^2 = m^2 - 4n_0 \in \mathbb{Z}.$$

Write $2b = u/v$ in lowest terms with $v > 0$. The last relation gives $v^2 \mid 5u^2$, so coprimality gives $v^2 \mid 5$. Hence $v = 1$, and $n = 2b$ is an integer. Furthermore $m^2 - 5n^2 = 4n_0$ implies $m^2 \equiv n^2 \pmod 4$. Since an integer square is 0 or 1 modulo four, m and n have the same parity. It follows that

$$\alpha = \frac{m + n\sqrt{5}}{2} = \frac{m - n}{2} + n\omega \in \mathbb{Z}[\omega].$$

Every element of $\mathbb{Z}[\omega]$ is integral, since ω is integral and integral elements form a ring. This finishes the computation. Normalization has added a missing element of the same fraction field; it has not enlarged the fraction field itself.

10.8 Existence of valuation rings

Recall that a valuation ring of a field K is a subring V such that, for each $0 \neq x \in K$, at least one of x, x^{-1} belongs to V. This property compares fractions directly and does not require a numerical valuation with values in $\mathbb{Z}$.

Proposition 10.30 (Three properties of valuation rings; AM 5.18)**.** *If V is a valuation ring of K, then:*

(i) V is a local ring;

(ii) every intermediate ring $V \subseteq V' \subseteq K$ is a valuation ring of K;

(iii) V is integrally closed in K.

Proof. For (i), let $\mathfrak{m}$ be the set of nonunits of V. If $x \in \mathfrak{m}$ and $a \in V$, then ax cannot be a unit, since an inverse of ax would give an inverse $a(ax)^{-1}$ of x. For nonzero $x, y \in \mathfrak{m}$, either $x/y \in V$ or $y/x \in V$. In the first case $x + y = y(1 + x/y) \in \mathfrak{m}$; the second is symmetric. Negatives and zero also belong to $\mathfrak{m}$. Thus $\mathfrak{m}$ is a proper ideal containing all nonunits, and every proper ideal is contained in it. It is the unique maximal ideal.

For (ii), the either-x-or-x^{-1} property is inherited by a larger subring of K.

For (iii), let $x \in K$ satisfy a monic equation over V. If $x \notin V$, then $x^{-1} \in V$. Rearranging the equation gives

$$x = -a_1 - a_2 x^{-1} - \cdots - a_n x^{-(n-1)} \in V,$$

a contradiction. Therefore $x \in V$. □

We next prove existence using Zorn's lemma. Fix a field K, an algebraically closed field Ω, and a homomorphism $f : A \to \Omega$ from a subring $A \subseteq K$. Consider all extensions (B, g), where $A \subseteq B \subseteq K$ and $g : B \to \Omega$ restricts to f. Order them by enlargement of the ring and extension of the map. A chain has an upper bound: take the union of its rings and use the compatible maps on that union. Zorn's lemma therefore supplies a maximal pair (B, g).

Lemma 10.31 (The kernel of a maximal extension; AM 5.19)**.** *For such a maximal pair (B, g), the ring B is local and its maximal ideal is $\mathfrak{m} = \ker g$.*

Proof. Since Ω is a field, $\ker g$ is prime. The map extends to $B_{\mathfrak{m}} \to \Omega$ by $b/s \mapsto g(b)/g(s)$, because $s \notin \mathfrak{m}$ makes the denominator nonzero. The domain B and its localization both lie in K. Maximality forces $B = B_{\mathfrak{m}}$, so the unique maximal ideal of this localization is already the unique maximal ideal of B. □

Lemma 10.32 (One of two extensions has a proper extended ideal; AM 5.20)**.** *For every $0 \neq x \in K$, at least one of $\mathfrak{m}B[x]$ and $\mathfrak{m}B[x^{-1}]$ is a proper ideal of the indicated ring.*

Proof. Suppose both equal their entire rings. Choose equations

$$1 = u_0 + u_1 x + \cdots + u_r x^r, \qquad 1 = v_0 + v_1 x^{-1} + \cdots + v_s x^{-s},$$

with all coefficients in $\mathfrak{m}$ and each degree minimal for its equation. Neither degree is zero, since $1 \notin \mathfrak{m}$. By swapping x and x^{-1} if necessary, assume $r \geq s$. Multiplication of the second equation by x^s gives

$$x^s = w_1 x^{s-1} + \cdots + w_s, \qquad w_i = (1-v_0)^{-1} v_i \in \mathfrak{m},$$

because $1 - v_0$ is a unit in the local ring. Replacing the term $u_r x^r$ in the first equation using this identity reduces its degree below r. All new coefficients still belong to $\mathfrak{m}$. This contradicts minimality of r. □

Theorem 10.33 (Maximal extensions are valuation rings; AM 5.21)**.** *The ring B in every maximal pair above is a valuation ring of K. In particular any $f : A \to \Omega$ extends to a homomorphism from some valuation ring of K containing A.*

Proof. Fix $0 \neq x \in K$. By Lemma 10.32, after possibly replacing x by x^{-1}, the ideal $\mathfrak{m}B[x]$ is proper. Choose a maximal ideal $\mathfrak{n}$ of $B[x]$ containing it. Its contraction to B is a proper ideal containing the maximal ideal $\mathfrak{m}$, so is $\mathfrak{m}$.

Write $k = B/\mathfrak{m}$ and $k' = B[x]/\mathfrak{n}$. The first is a subfield of the second, and $k' = k[\bar{x}]$. The element $\bar{x}$ must be algebraic over k: otherwise $k[\bar{x}]$ would be a polynomial ring, in which $\bar{x}$ has no inverse, and would not be a field. Consequently k'/k is a finite algebraic extension generated by $\bar{x}$.

The map g embeds k in Ω. A minimal polynomial for $\bar{x}$ has a root in Ω; sending $\bar{x}$ to that root extends the embedding to k'. Compose $B[x] \to k' \to \Omega$ to extend g. Maximality of (B, g) forces $B[x] = B$, so $x \in B$. Thus for each original fraction either it or its inverse belongs to B. □

Corollary 10.34 (Integral closure as an intersection; AM 5.22)**.** *For a subring A of a field K, its integral closure in K is the intersection of all valuation rings of K containing A.*

Proof. Every such valuation ring is integrally closed, so contains the integral closure of A.

Conversely let $x \in K$ not be integral over A; necessarily $x \neq 0$. Put $A' = A[x^{-1}]$. The element x^{-1} is not a unit in A'. Indeed, if $x \in A'$, write $x = a_0 + a_1 x^{-1} + \cdots + a_r x^{-r}$ and multiply by x^r to obtain a monic equation for x over A, a contradiction.

Choose a maximal ideal $\mathfrak{n}$ of A' containing x^{-1}, and embed $A'/\mathfrak{n}$ in its algebraic closure Ω. The resulting map $A' \to \Omega$ extends, by Theorem 10.33, to a valuation ring $V \supseteq A'$. The extended homomorphism sends x^{-1} to zero. Thus x cannot belong to V, since otherwise the homomorphism would send $xx^{-1} = 1$ to zero. We have found a valuation ring containing A but excluding x. □

Proposition 10.35 (Extending maps while avoiding a zero; AM 5.23)**.** *Let $A \subseteq B$ be domains, with B finitely generated as an A-algebra, and let $0 \neq v \in B$. There exists $0 \neq u \in A$ such that every homomorphism $f : A \to \Omega$ to an algebraically closed field with $f(u) \neq 0$ extends to $g : B \to \Omega$ satisfying $g(v) \neq 0$.*

Proof. First suppose $B = A[x]$. If x is transcendental over $\mathrm{Frac}(A)$, write $v = a_d x^d + \cdots + a_0$ with $a_d \neq 0$ and take $u = a_d$. The polynomial $f(a_d)T^d + \cdots + f(a_0)$ is nonzero. An algebraically closed field is infinite, so it has an element c that is not a root. Evaluation $x \mapsto c$ defines the required g. For completeness, a finite field with elements $c_1, \ldots, c_N$ cannot be algebraically closed because $\prod_i (T - c_i) - 1$ has no root in it.

If x is algebraic over $\mathrm{Frac}(A)$, so is v^{-1}. Clear denominators in equations for these two elements. Let $u \in A \setminus \{0\}$ be the product of their nonzero leading coefficients. After

inverting u, both x and v^{-1} are integral over $A[1/u]$. The map f extends to $A[1/u]$ because $f(u) \neq 0$. By Theorem 10.33, it extends further to a valuation ring V of $\mathrm{Frac}(B)$ containing $A[1/u]$. Since V is integrally closed, it contains x and v^{-1}, hence also B. The image of v is nonzero because v is a unit of V. Restrict the extended map to B.

For several generators, write $B = A'[x]$ with A' generated by one fewer element. The one-generator case supplies $0 \neq u' \in A'$ such that any map on A' avoiding u' extends to B avoiding v. By induction there is $0 \neq u \in A$ such that a map avoiding u extends to A' avoiding u'. Composing the two extensions proves the statement. □

Corollary 10.36 (The field form of the Nullstellensatz; AM 5.24)**.** *If a field L is finitely generated as an algebra over a field k, then L/k is a finite algebraic extension.*

Proof. Apply Proposition 10.35 with $A = k$, $B = L$, $v = 1$, and the inclusion $k \to \overline{k}$. Its nonzero element $u \in k$ automatically has nonzero image. We obtain a homomorphism $L \to \overline{k}$, which is injective because L is a field. Thus all the algebra generators of L are algebraic over k. Adjoining finitely many algebraic elements gives a finite extension: the powers below their successive minimal-polynomial degrees span it. □

For example $k(t)$ is generated by t as a *field*, but it cannot be finitely generated as a *k-algebra*. A finite list of rational functions has only finitely many denominator factors; inverses of new irreducible factors are missing. Corollary 10.36 abstracts precisely this obstruction.

10.9 Base change and finite projections

For A-algebras B and C, replacing B by $B \otimes_A C$ is called *base change from A to C*. The new C-algebra structure is $c \mapsto 1 \otimes c$. An exercise at the end of this lecture asks whether integrality survives this operation.

For the geometric interpretation, recall the coordinate-ring construction. If $X \subseteq k^n$ is an affine algebraic set, then

$$I(X) = \{f \in k[T_1, \ldots, T_n] : f(a) = 0 \text{ for every } a \in X\}, \qquad k[X] = k[T_1, \ldots, T_n]/I(X).$$

Its elements are the polynomial functions restricted to X. A k-algebra homomorphism $k[X] \to k$ determines numbers a_i by the images of the coordinate classes. All equations in $I(X)$ vanish at this tuple. Since X is itself the common zero set of polynomial equations, this forces $(a_1, \ldots, a_n) \in X$. Conversely evaluation at any point gives such a homomorphism. We may therefore pass between points and homomorphisms to k in the following argument.

Theorem 10.37 (Noether normalization; AM Chapter 5, Exercise 16)**.** *Let $0 \neq A = k[x_1, \ldots, x_n]$ be a finitely generated algebra over any field k. There are algebraically independent elements $y_1, \ldots, y_r \in A$ such that A is finite over $k[y_1, \ldots, y_r]$. If k is infinite, the y_i may be chosen as k-linear combinations of the original generators. If k is algebraically closed and $A = k[X]$ for a nonempty affine algebraic set $X \subseteq k^n$, these linear forms give a surjective map $X \to k^r$.*

Proof. We first give a proof valid even over finite fields. Induct on the number n of generators. If the generators are algebraically independent, use them as the y_i; this includes $n = 0$. Otherwise choose a nonzero relation $F(x_1, \ldots, x_n) = 0$ with $F \in k[T_1, \ldots, T_n]$.

Choose an integer M larger than every exponent appearing in any monomial of F. Set $w_i = M^i$ for $i < n$, and set

$$z_i = x_i - x_n^{w_i} \quad (1 \leq i < n).$$

After substituting $T_i = Z_i + T_n^{w_i}$ into F, a monomial $T_1^{a_1} \cdots T_n^{a_n}$ has highest T_n-degree

$$a_n + a_1 M + \cdots + a_{n-1} M^{n-1}.$$

These integers are distinct for distinct monomials, by uniqueness of base-M expansion. Hence there is a unique largest degree, and its coefficient is the nonzero coefficient in k of that monomial. Every term involving one of the Z_i has lower T_n-degree. Dividing by the nonzero leading scalar makes the relation a monic equation for x_n over $A' = k[z_1, \ldots, z_{n-1}]$.

Thus $A = A'[x_n]$ is finite over A', since every original x_i equals $z_i + x_n^{w_i}$. By induction, A' is finite over a polynomial subalgebra $k[y_1, \ldots, y_r]$. Products of finite generating lists show that A is finite over that same subalgebra. This proves the assertion for every field, without assuming A is a domain.

Now suppose k is infinite. Replace the powers above by a linear substitution. The *total degree* of a monomial $T_1^{a_1} \cdots T_n^{a_n}$ is $a_1 + \cdots + a_n$. A polynomial is *homogeneous* if all its nonzero monomials have the same total degree. Grouping terms of equal total degree writes F uniquely as a sum of homogeneous parts. Let H be its nonzero part of highest total degree. The polynomial $H(U_1, \ldots, U_{n-1}, 1)$ is nonzero: its different monomials remain different after setting the last variable to 1, by homogeneity. A nonzero polynomial over an infinite field is nonzero at some point. To prove this by induction on the number of variables, write it as a polynomial in its last variable and choose a nonzero coefficient polynomial in the other variables. By induction there are values of those variables at which this coefficient is nonzero. The resulting nonzero one-variable polynomial has finitely many roots, so choose the last coordinate outside that finite set. Choose $c_i \in k$ with $H(c_1, \ldots, c_{n-1}, 1) \neq 0$, and set $z_i = x_i - c_i x_n$. The relation after this substitution has a nonzero constant leading coefficient in x_n, so again x_n is integral over A'. Repeating inductively yields y_j that are linear combinations of the original x_i.

If k is algebraically closed and A is the coordinate ring of a nonempty affine algebraic set $X \subseteq k^n$, the functions y_j define a linear map $X \to k^r$. Given any point $(a_1, \ldots, a_r) \in k^r$, let $\mathfrak{n} = (y_1 - a_1, \ldots, y_r - a_r)$ in $B = k[y_1, \ldots, y_r]$. Its quotient is k, so $\mathfrak{n}$ is maximal. Lying over (Theorem 10.16) supplies a prime $\mathfrak{q} \subseteq A$ with $\mathfrak{q} \cap B = \mathfrak{n}$. The domain $A/\mathfrak{q}$ is integral over $B/\mathfrak{n} = k$, hence is a field by Proposition 10.13; every one of its elements is algebraic over k. Since k is algebraically closed, each such element lies in k, so the quotient map gives a k-algebra homomorphism $A \to k$. The point–homomorphism correspondence above supplies a point of X mapping to the chosen tuple. Thus the projection is surjective. Algebraic independence of the linear forms makes the ambient linear map $k^n \to k^r$ surjective as well, and choosing a preimage in k^n of each standard basis vector gives an injective linear map $k^r \to k^n$ whose composition with the projection is the identity. Its image is an r-dimensional linear subspace.

When A is a domain, its fraction field is a finite algebraic extension of $k(y_1,\dots,y_r)$. Indeed the finitely many algebra generators of A are integral, hence algebraic, over this field, and adjoining them produces the fraction field of A. This description prepares for the dimension calculation in Lecture 12. □

Example 10.38 (Normalization need not preserve a chosen projection)**.** For the cusp $A = k[t^2,t^3]$, normalization is the inclusion $A \subseteq k[t]$. Noether normalization is a different operation: $k[t^2] \subseteq A$ is a finite extension of a polynomial subring, with $A = k[t^2] + k[t^2]t^3$. One enlarges the ring to remove missing integral fractions; the other finds polynomial coordinates over which the given ring is finite.

10.10 Recognizing a discrete valuation ring

A dimension-one domain has only two kinds of prime ideals: the zero ideal and nonzero maximal ideals. This condition is easy to visualize for $k[t]$: the zero ideal describes the entire line, while a nonzero prime records an irreducible polynomial. Over an algebraically closed field these are the points $(t-a)$.

We need two elementary observations before giving the recognition theorem. First, if $(A,\mathfrak{m})$ is a Noetherian local domain of dimension one and $0 \neq a \in \mathfrak{m}$, the only prime containing (a) is $\mathfrak{m}$. Thus $\sqrt{(a)} = \mathfrak{m}$. Choose generators $u_1,\dots,u_s$ of $\mathfrak{m}$ and positive integers e_i with $u_i^{e_i} \in (a)$. Every monomial in the generators of degree

$$N = 1 + \sum_i (e_i - 1)$$

contains one of the powers $u_i^{e_i}$, so $\mathfrak{m}^N \subseteq (a)$. This elementary counting argument is often called the pigeonhole argument for powers of an ideal.

Second, a Noetherian local domain with a nonzero principal maximal ideal $\mathfrak{m} = (\pi)$ is already a DVR. We can construct the valuation directly. If $0 \neq a$ were divisible by every π^n, the elements $b_n = a/\pi^n$ would all lie in A. Their principal ideals form an ascending chain because $b_n = \pi b_{n+1}$. When the chain stabilizes, $b_{n+1} = cb_n$ for some $c \in A$, so cancellation gives $1 = c\pi$, a contradiction. Hence each nonzero a has a largest exponent r for which $a \in (\pi^r)$. Write $a = \pi^r u$. Maximality of r says $u \notin \mathfrak{m}$, so u is a unit. Products add these exponents, and sums have at least the smaller exponent. Extending by $v(a/b) = v(a) - v(b)$ gives a discrete valuation. In particular every nonzero ideal is a power of (π), by choosing an element of least exponent.

Proposition 10.39 (Six tests for a DVR; AM 9.2)**.** *Let $(A,\mathfrak{m})$ be a Noetherian local domain of dimension one, with residue field $k = A/\mathfrak{m}$. The following conditions are equivalent:*

(i) A is a discrete valuation ring.

(ii) A is integrally closed.

(iii) The maximal ideal $\mathfrak{m}$ is principal.

(iv) $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = 1$.

(v) Every nonzero ideal is a power of $\mathfrak{m}$, including $A = \mathfrak{m}^0$.

(vi) There is an element $\pi \in A$ such that every nonzero ideal has the form (π^r) for some integer $r \geq 0$.

Proof. We give enough implications to connect all six conditions, recording exactly where finite generation is used.

(i) implies (ii) and (v). A DVR is a valuation ring, hence integrally closed by Proposition 10.30. The least-valuation argument at the start of this lecture proves (v).

(ii) implies (iii). Fix $0 \neq a \in \mathfrak{m}$. By the first observation above there is a least $N \geq 1$ with $\mathfrak{m}^N \subseteq (a)$. Choose $b \in \mathfrak{m}^{N-1} \setminus (a)$ and put $c = b/a \in \mathrm{Frac}(A)$. Then $c \notin A$, but $c\mathfrak{m} \subseteq A$. If $c\mathfrak{m} \subseteq \mathfrak{m}$, multiplication by c preserves the finitely generated module $\mathfrak{m}$. The determinant argument in Proposition 10.4 supplies a monic polynomial P over A with $P(c)\mathfrak{m} = 0$. Since $\mathfrak{m} \neq 0$ and A is a domain, this gives $P(c) = 0$ in its fraction field. Normality would imply $c \in A$, a contradiction. Thus the ideal $c\mathfrak{m}$ of A contains an element outside $\mathfrak{m}$, which is a unit. Therefore $c\mathfrak{m} = A$, and multiplying by c^{-1} yields $\mathfrak{m} = c^{-1}A$.

(iii) implies (i) and (vi). Apply the second observation above, which constructs the valuation explicitly and describes all ideals.

(iii) implies (iv). A generator of $\mathfrak{m}$ spans $\mathfrak{m}/\mathfrak{m}^2$. This vector space is nonzero: otherwise $\mathfrak{m} = \mathfrak{m}^2$, and Nakayama's lemma applied to the finite module $\mathfrak{m}$ gives $\mathfrak{m} = 0$, contradicting dimension one. Thus its dimension is exactly one.

(iv) implies (iii). Lift a basis vector of $\mathfrak{m}/\mathfrak{m}^2$ to $\pi \in \mathfrak{m}$. Then $\mathfrak{m} = (\pi) + \mathfrak{m}^2$. The quotient module $M = \mathfrak{m}/(\pi)$ satisfies $M = \mathfrak{m}M$. It is finite, so Nakayama gives $M = 0$ and $\mathfrak{m} = (\pi)$.

(v) implies (iii). Again $\mathfrak{m} \neq \mathfrak{m}^2$. Choose $\pi \in \mathfrak{m} \setminus \mathfrak{m}^2$. By (v), $(\pi) = \mathfrak{m}^r$ for some $r \geq 1$. If $r \geq 2$, then $\pi \in \mathfrak{m}^2$, a contradiction. Hence $r = 1$.

(vi) implies (iii). The nonzero proper ideal $\mathfrak{m}$ is (π^r) for some $r \geq 1$. The element π is nonzero and is not a unit, so

$$(\pi) \subseteq \mathfrak{m} = (\pi^r) \subseteq (\pi).$$

All three ideals are equal. This completes the equivalence. □

Example 10.40 (The cusp fails the DVR test)**.** At the origin of the cusp, let

$$R = k[t^2, t^3]_{(t^2,t^3)}, \qquad \mathfrak{m} = (t^2, t^3)R.$$

The classes of t^2, t^3 form a basis of $\mathfrak{m}/\mathfrak{m}^2$ over k. They span because they generate $\mathfrak{m}$. For independence, a relation $at^2 + bt^3 \in \mathfrak{m}^2$ with $a, b \in k$ can be cleared by a denominator whose constant term is nonzero. But elements of $(t^2, t^3)^2$ have no terms of degrees 2 or 3. Comparing those coefficients forces $a = b = 0$. The vector space $\mathfrak{m}/\mathfrak{m}^2$, called the *cotangent space* of this local ring, therefore has dimension two. This prevents R from being a DVR, since a DVR has a principal maximal ideal. The element t gives a second way to detect the same failure. It is integral over R because $t^2 \in R$, but it does not belong to R: an equation $t = a/b$ with $a, b \in k[t^2, t^3]$ and $b(0) \neq 0$ would make $a = tb$ have a nonzero coefficient of t. No element of $k[t^2, t^3]$ has such a coefficient. Thus R is not integrally closed. For comparison, the local ring $k[t]_{(t)}$ of the line has one-dimensional cotangent space.

10.11 Dedekind domains and ideal factorization

We will use the word *primary* only for an individual ideal in the next criterion. A proper ideal Q is primary if $ab \in Q$ and $a \notin Q$ imply $b^n \in Q$ for some $n \geq 1$. Equivalently, every zero divisor in A/Q is nilpotent. If $\sqrt{Q}$ is a maximal ideal $\mathfrak{p}$, then Q is primary: the quotient A/Q has the unique maximal ideal $\mathfrak{p}/Q$, all its elements in that ideal are nilpotent, and every element outside it is a unit. A zero divisor cannot be a unit. This proof does not require any decomposition theorem.

Theorem 10.41 (The local criterion for a Dedekind domain; AM 9.3)**.** *For a Noetherian domain A of dimension one, the following are equivalent:*

(i) A is integrally closed.

(ii) Every primary ideal is a power of a prime ideal.

(iii) $A_\mathfrak{p}$ is a discrete valuation ring for every nonzero prime ideal $\mathfrak{p}$.

A domain satisfying these conditions is called a Dedekind domain*. We use the dimension-one convention, so a field is not called a Dedekind domain here.*

Proof. For (i)⇔(iii), use the local criterion for normality, Proposition 10.20, and the DVR criterion, Proposition 10.39. The localization at the zero prime is the fraction field, which is already integrally closed. For $\mathfrak{p} \neq 0$, the primes of $A_\mathfrak{p}$ correspond to (0) and $\mathfrak{p}$, so this local ring is a Noetherian domain of dimension one.

Assume (iii), and let Q be primary. If $Q = 0$, it is the first power of the zero prime, so suppose $Q \neq 0$. Its radical $\mathfrak{p}$ is prime: if $(ab)^n \in Q$ and $a \notin \sqrt{Q}$, primaryness forces some power of b^n into Q. Since $\mathfrak{p}$ is nonzero, it is maximal. In the DVR $A_\mathfrak{p}$ write $QA_\mathfrak{p} = (\mathfrak{p}A_\mathfrak{p})^r$ with $r \geq 1$. We claim that contracting this equality gives $Q = \mathfrak{p}^r$. Indeed, Q is unchanged by extension and contraction at $\mathfrak{p}$: if $sx \in Q$ for $s \notin \mathfrak{p}$, primaryness forces $x \in Q$. The same is true of $\mathfrak{p}^r$, since its radical is the maximal ideal $\mathfrak{p}$ and hence it is primary by the preceding observation. This proves (ii).

Conversely assume (ii), fix a nonzero prime $\mathfrak{p}$, and let J be a nonzero proper ideal of $A_\mathfrak{p}$. This local ring is one dimensional, so the radical of J is its maximal ideal. Hence J is primary by the same elementary observation. The contracted ideal $Q = J \cap A$ is primary: apply the defining implication after mapping A into $A_\mathfrak{p}$. Its radical is $\mathfrak{p}$. By (ii), $Q = \mathfrak{p}^r$, so extension gives $J = (\mathfrak{p}A_\mathfrak{p})^r$. Here every ideal of a localization is the extension of its contraction. Condition (v) of Proposition 10.39 now says $A_\mathfrak{p}$ is a DVR. □

The local DVR description gives a direct proof of the following factorization theorem for ideals.

Corollary 10.42 (Unique prime factorization of ideals; AM 9.4)**.** *Every nonzero proper ideal I in a Dedekind domain A has a unique expression*

$$I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r},$$

where the $\mathfrak{p}_i$ are distinct nonzero prime ideals and the e_i are positive integers. The unit ideal is represented by the empty product.

Proof. Every prime containing $I \neq 0$ is maximal. Thus A/I is Noetherian of dimension zero, hence Artinian, and there are only finitely many maximal ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ containing I.

Each $A_{\mathfrak{p}_i}$ is a DVR. Define $e_i \geq 1$ by

$$IA_{\mathfrak{p}_i} = (\mathfrak{p}_i A_{\mathfrak{p}_i})^{e_i}, \qquad J = \prod_{i=1}^{r} \mathfrak{p}_i^{e_i}.$$

Distinct maximal ideals become unit ideals after localization at each other. Hence, for every maximal ideal $\mathfrak{m}$,

$$IA_{\mathfrak{m}} = JA_{\mathfrak{m}} = \begin{cases} (\mathfrak{p}_i A_{\mathfrak{p}_i})^{e_i}, & \mathfrak{m} = \mathfrak{p}_i, \\ A_{\mathfrak{m}}, & \mathfrak{m} \notin \{\mathfrak{p}_1, \dots, \mathfrak{p}_r\}. \end{cases}$$

To deduce $I = J$, suppose $x \in I \setminus J$. The ideal $(J : x)$ is proper, so choose a maximal ideal $\mathfrak{m} \supseteq (J : x)$. Then

$$x/1 \in JA_{\mathfrak{m}} \quad \Longrightarrow \quad sx \in J \text{ for some } s \notin \mathfrak{m} \quad \Longrightarrow \quad s \in (J : x) \subseteq \mathfrak{m},$$

a contradiction. Thus $I \subseteq J$; interchanging I, J gives equality.

For uniqueness, localize any factorization at a nonzero prime $\mathfrak{p}$. Every other factor becomes the unit ideal. The remaining exponent is the uniquely determined integer e in

$$IA_{\mathfrak{p}} = (\mathfrak{p}A_{\mathfrak{p}})^{e}.$$

Thus all prime exponents are uniquely determined. □

Example 10.43 (A familiar factorization and its meaning)**.** For $A = \mathbb{Z}$,

$$(360) = (2)^3(3)^2(5), \qquad (360)\mathbb{Z}_{(2)} = (2^3)\mathbb{Z}_{(2)}.$$

The factors $3, 5$ are units in $\mathbb{Z}_{(2)}$. In $\mathbb{Z}[i]$,

$$5 = (2+i)(2-i), \qquad (5) = (2+i)\,(2-i).$$

Here the two factors on the right in the second equality are ideals. Ideal factorization remains unique even in Dedekind domains without unique factorization of elements.

A finite extension $K/\mathbb{Q}$ is called a *number field*. An element of K integral over $\mathbb{Z}$ is an *algebraic integer*. Thus the ring of algebraic integers in K is exactly the integral closure of $\mathbb{Z}$ in K.

Theorem 10.44 (Rings of algebraic integers; AM 9.5)**.** *Let K be a finite extension of $\mathbb{Q}$, and let A be the integral closure of $\mathbb{Z}$ in K. Then A is a Dedekind domain.*

Proof. In characteristic zero, an irreducible polynomial f has $f' \neq 0$ and $\deg f' < \deg f$, so $\gcd(f, f') = 1$. Hence $K/\mathbb{Q}$ is separable. Proposition 10.29 gives a $\mathbb{Q}$-basis $v_1, \dots, v_n$ with

$$A \subseteq \sum_{i=1}^{n} \mathbb{Z} v_i \cong \mathbb{Z}^n.$$

Since $\mathbb{Z}$ is Noetherian, A is a finite $\mathbb{Z}$-module. For every ideal $I \subseteq A$, finite generation over $\mathbb{Z}$ gives

$$I = \sum_{j=1}^{r} \mathbb{Z} a_j \implies I = \sum_{j=1}^{r} A a_j.$$

Thus A is Noetherian.

For $x \in K$, choose a monic equation

$$x^m + c_1 x^{m-1} + \cdots + c_m = 0, \qquad c_i \in \mathbb{Q}.$$

Choose $0 \neq d \in \mathbb{Z}$ with $dc_i \in \mathbb{Z}$ for all i. Multiplication by d^m gives

$$(dx)^m + (dc_1)(dx)^{m-1} + \cdots + d^m c_m = 0,$$

whose coefficients are integers. Therefore $dx \in A$ and $x = (dx)/d \in \mathrm{Frac}(A)$; hence $\mathrm{Frac}(A) = K$. Transitivity of integrality gives

$$z \in K \text{ integral over } A \implies z \text{ integral over } \mathbb{Z} \implies z \in A.$$

So A is normal.

For a nonzero prime $\mathfrak{p} \subset A$, incomparability excludes

$$(0) \subsetneq \mathfrak{p}, \qquad (0) \cap \mathbb{Z} = \mathfrak{p} \cap \mathbb{Z}.$$

Thus $\mathfrak{p} \cap \mathbb{Z}$ is a nonzero prime, hence maximal. Proposition 10.14 makes $\mathfrak{p}$ maximal. Finally A is not a field: by Proposition 10.13, the integral inclusion $\mathbb{Z} \subseteq A$ would otherwise make $\mathbb{Z}$ a field. It follows that $\dim A = 1$. The ring A is therefore Dedekind. □

10.12 Fractional ideals and their inverses

A fractional ideal of a domain A is an A-submodule $M \subseteq K = \mathrm{Frac}(A)$ for which $dM \subseteq A$ for some $0 \neq d \in A$. An ordinary ideal is a fractional ideal with $d = 1$. For example $(1/6)\mathbb{Z}$ is a fractional ideal of $\mathbb{Z}$; it is a set of fractions with a common denominator. Every finitely generated A-submodule of K is fractional, because multiplying the denominators of a finite generating list clears all of them. If A is Noetherian, the converse holds: dM is an ordinary ideal and thus finitely generated.

The product MN consists of finite sums of products xy with $x \in M$, $y \in N$. Call M *invertible* if $MN = A$ for some submodule $N \subseteq K$. Its inverse must be

$$(A : M) = \{y \in K : yM \subseteq A\}.$$

Indeed $N \subseteq (A : M)$, while

$$(A : M) = (A : M)MN \subseteq AN = N.$$

Invertibility also forces finite generation: write $1 = \sum_{i=1}^{r} x_i y_i$ with $x_i \in M$ and $y_i \in (A : M)$. For every $x \in M$,

$$x = \sum_i (xy_i)x_i,$$

whose coefficients lie in A. Thus the x_i generate M. In particular the definition using arbitrary submodules produces fractional ideals automatically.

Proposition 10.45 (Invertibility is local; AM 9.6)**.** *For a fractional ideal M of a domain A, the following are equivalent:*

(i) M is invertible.

(ii) M is finitely generated and $M_{\mathfrak{p}}$ is invertible over $A_{\mathfrak{p}}$ for every prime $\mathfrak{p}$.

(iii) M is finitely generated and $M_{\mathfrak{m}}$ is invertible over $A_{\mathfrak{m}}$ for every maximal ideal $\mathfrak{m}$.

Proof. If M is invertible, it is finite by the preceding argument. Localization of $M(A : M) = A$ shows that each $M_{\mathfrak{p}}$ is invertible. Thus (i) implies (ii), and (ii) implies (iii).

For the converse, finite generation gives

$$(A : M)_{\mathfrak{m}} = (A_{\mathfrak{m}} : M_{\mathfrak{m}})$$

as submodules of K. To check the only nontrivial containment, suppose $z \in K$ satisfies $zM \subseteq A_{\mathfrak{m}}$. For generators $x_1, \ldots, x_r$ of M, choose $s_i \notin \mathfrak{m}$ with $s_i z x_i \in A$. Their product s then satisfies $szM \subseteq A$, so $z \in (A : M)_{\mathfrak{m}}$. This is the precise place where a finite generating list is needed.

Set $I = M(A : M) \subseteq A$. Under (iii), its localization at every maximal ideal is

$$I_{\mathfrak{m}} = M_{\mathfrak{m}}(A_{\mathfrak{m}} : M_{\mathfrak{m}}) = A_{\mathfrak{m}}.$$

A proper ideal would lie in a maximal ideal and remain proper there. Hence $I = A$, proving invertibility. The zero module satisfies none of these equivalent conditions in a domain, so no separate nonzero hypothesis was necessary. □

Proposition 10.46 (The local inverse-ideal test; AM 9.7)**.** *Let A be a local domain which is not a field. Then A is a DVR if and only if every nonzero fractional ideal of A is invertible.*

Proof. In a DVR with uniformizer π, a nonzero fractional ideal M becomes an ordinary ideal after multiplication by $d \neq 0$. Since ordinary nonzero ideals are (π^r), we get $M = (\pi^{r - v(d)})$. Its inverse is generated by the opposite power of π.

Conversely assume all nonzero fractional ideals are invertible. Every nonzero ordinary ideal is then finitely generated, so A is Noetherian. We claim that an invertible ideal in a local ring is principal. Write $1 = \sum_i x_i y_i$ as above. At least one summand $x_i y_i$ is a unit, since otherwise the sum would lie in the maximal ideal. Fix that index. For any $x \in M$,

$$\frac{x}{x_i} = \frac{xy_i}{x_i y_i} \in A.$$

Thus $M = Ax_i$. Apply this to the nonzero maximal ideal $\mathfrak{m}$ of A. It is principal, and the ascending-chain argument preceding Proposition 10.39 constructs a discrete valuation. Hence A is a DVR. □

Theorem 10.47 (The global inverse-ideal test; AM 9.8)**.** *For a domain A, every nonzero fractional ideal is invertible if and only if A is either a field or a Dedekind domain.*

Proof. A field has only one nonzero submodule of itself, namely itself. Now suppose A is Dedekind and $M \neq 0$ is fractional. It is finite because A is Noetherian. At each nonzero prime, its localization is invertible by Proposition 10.46; at the zero prime it is the fraction field. Proposition 10.45 makes M invertible.

Conversely assume every nonzero fractional ideal is invertible and that A is not a field. Ordinary ideals are finite, so A is Noetherian. For any maximal ideal $\mathfrak{m}$, every nonzero ideal of $A_\mathfrak{m}$ is the extension of a nonzero ideal of A. The latter is invertible, and its extension is invertible. Every fractional ideal of $A_\mathfrak{m}$ is a nonzero scalar multiple of an ordinary ideal, so it too is invertible. Since $\mathfrak{m} \neq 0$, the local ring $A_\mathfrak{m}$ is not a field. Proposition 10.46 makes it a DVR.

It follows that A is integrally closed by Proposition 10.20. If $\mathfrak{p} \neq 0$ is any prime, choose a maximal ideal $\mathfrak{m} \supseteq \mathfrak{p}$. Its extension in $A_\mathfrak{m}$ is a nonzero prime, hence the maximal ideal of that DVR. Contracting gives $\mathfrak{p} = \mathfrak{m}$. Thus every nonzero prime is maximal, and A has dimension one. It is Dedekind. □

Remark 10.48 (The field boundary in AM 9.7–9.8). The displayed formulations in the book do not state the field exception explicitly. With its definitions of a DVR by a surjective integer-valued valuation and a Dedekind domain by dimension one, a field satisfies the invertibility condition but neither definition. We have stated the nonfield hypothesis in Proposition 10.46 and included fields in Theorem 10.47, making the boundary case explicit.

Corollary 10.49 (The group of fractional ideals; AM 9.9)**.** *The nonzero fractional ideals of a Dedekind domain form an abelian group under multiplication, with identity A and inverse $M^{-1} = (A : M)$.*

Proof. Theorem 10.47 gives inverses. The product of two fractional ideals is fractional: if $dM \subseteq A$ and $eN \subseteq A$, then $deMN \subseteq A$. Products are nonzero in a domain. Associativity and commutativity follow by expanding finite sums of products, and $AM = M$. These verify all the group axioms. □

The *free abelian group on a set* consists of finite integer linear combinations of its elements. For a nonzero fractional ideal M, choose $0 \neq d \in A$ with $dM \subseteq A$. Factor the two ordinary ideals:

$$dM = \prod_\mathfrak{p} \mathfrak{p}^{a_\mathfrak{p}}, \qquad (d) = \prod_\mathfrak{p} \mathfrak{p}^{b_\mathfrak{p}}.$$

Only finitely many exponents are nonzero. Invertibility then gives

$$M = (d)^{-1}(dM) = \prod_\mathfrak{p} \mathfrak{p}^{a_\mathfrak{p} - b_\mathfrak{p}}.$$

Localization in each DVR determines these integer exponents uniquely, including negative exponents. Consequently the fractional-ideal group is

$$\mathscr{I}(A) \cong \bigoplus_{\substack{\mathfrak{p}\in \operatorname{Spec} A \\ \mathfrak{p}\neq(0)}} \mathbb{Z}.$$

A *principal fractional ideal* is $Ax = (x)$ for some $x \in K^\times$. These ideals form a subgroup, because $(x)(y) = (xy)$ and $(x)^{-1} = (x^{-1})$. The quotient is the *ideal class group* $\operatorname{Cl}(A)$, measuring which ideals fail to have a single generator. Exactness for a sequence of group homomorphisms means that the image of each homomorphism is the kernel of the next, just as for modules; the trivial multiplicative group is written 1. There is an exact sequence of abelian groups

$$1 \longrightarrow A^\times \longrightarrow K^\times \xrightarrow{x\mapsto(x)} \mathscr{I}(A) \longrightarrow \operatorname{Cl}(A) \longrightarrow 1,$$

where $\mathscr{I}(A)$ is the fractional-ideal group. The kernel of $x \mapsto (x)$ is exactly the units, because $(x) = A$ means both x and x^{-1} lie in A. The class group is zero exactly when every ideal is principal. This definition needs no finiteness theorem for class groups.

Checking an ideal factorization by generators

Let $A = \mathbb{Z}[\sqrt{-5}]$ and $\mathfrak{p} = (2, 1+\sqrt{-5})$. The quotient $A/\mathfrak{p} \cong \mathbf{F}_2$, so $\mathfrak{p}$ is maximal. Write $s = \sqrt{-5}$. The square of the ideal is generated by

$$4, \qquad 2+2s, \qquad (1+s)^2 = -4+2s.$$

All three generators lie in (2), so $\mathfrak{p}^2 \subseteq (2)$. Conversely,

$$(2+2s) - (-4+2s) = 6, \qquad 6-4 = 2,$$

so $2 \in \mathfrak{p}^2$ and $\mathfrak{p}^2 = (2)$. This is equality of ideals, established in both directions.

The ideal $\mathfrak{p}$ is not principal. If $\mathfrak{p} = (a+bs)$, its index in A would equal the absolute determinant of multiplication by $a+bs$ on the basis $1, s$:

$$\left|\det \begin{pmatrix} a & -5b \\ b & a \end{pmatrix}\right| = a^2 + 5b^2.$$

But $[A : \mathfrak{p}] = 2$, whereas $a^2 + 5b^2 = 2$ has no integer solution. Thus an ideal can have a principal square without itself being principal. The fractional ideal $\frac{1}{2}\mathfrak{p}$ is its inverse, since $\mathfrak{p}(\frac{1}{2}\mathfrak{p}) = A$. This example makes the difference between factoring elements and factoring ideals visible in an explicit calculation.

10.13 Exercises

Exercise 10.1. In $\mathbb{Z}_{(3)}$, simplify the ideal generated by 18 and 45, and compute the length of $\mathbb{Z}_{(3)}/(27)$.

Exercise 10.2. Find a monic polynomial with coefficients in $\mathbb{Z}$ that vanishes at $\sqrt{2}+\sqrt{3}$. Explain why $\mathbb{Z}[1/2]$ is of finite type but is not finite over $\mathbb{Z}$.

Exercise 10.3. Compute the conductor of $k[t^2,t^3] \subseteq k[t]$.

Exercise 10.4 (Integral maps are closed; AM Chapter 5, Exercise 1)**.** A map of topological spaces is *closed* if it sends every closed subset to a closed subset. For an integral homomorphism $f : A \to B$, prove that the contraction map $f^* : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is closed. Here $V_B(J)$ denotes the primes containing $J \subseteq B$, and these are the closed sets.

Exercise 10.5 (Extending a point through an integral extension; AM Chapter 5, Exercise 2)**.** Let $A \subseteq B$ be integral and let $f : A \to \Omega$ be a homomorphism to an algebraically closed field. Prove that f extends to $B \to \Omega$.

Exercise 10.6 (Integrality under base change; AM Chapter 5, Exercise 3)**.** If $f : B \to B'$ is an integral homomorphism of A-algebras and C is an A-algebra, prove that $f \otimes 1 : B \otimes_A C \to B' \otimes_A C$ is integral.

Exercise 10.7 (A localization trap; AM Chapter 5, Exercise 4)**.** For an integral inclusion $A \subseteq B$, a maximal ideal $\mathfrak{n}$ of B, and $\mathfrak{m} = \mathfrak{n} \cap A$, must $B_{\mathfrak{n}}$ be integral over $A_{\mathfrak{m}}$?

Exercise 10.8 (Finite symmetries give integral extensions; AM Chapter 5, Exercise 12)**.** Let a finite group G act on a ring A by automorphisms, and write $A^G = \{a : \sigma(a) = a \text{ for all } \sigma \in G\}$. Prove that A is integral over A^G. If a multiplicatively closed set $S \subseteq A$ satisfies $\sigma(S) \subseteq S$ for every σ, put $S^G = S \cap A^G$. Prove that the action extends to $S^{-1}A$ and that

$$(S^G)^{-1}(A^G) \cong (S^{-1}A)^G.$$

Exercise 10.9 (A finite projection of a hyperbola)**.** Let $A = k[t,t^{-1}]$ and put $u = t + t^{-1}$. Prove that u is algebraically independent over k and that A is a free $k[u]$-module with basis $1, t$. If k is algebraically closed, prove that every value $a \in k$ is attained by $t + t^{-1}$ with $t \neq 0$, and that at most two values of t give a prescribed a.

Exercise 10.10 (Content and Gauss's lemma; AM Chapter 9, Exercise 2)**.** Let A be a Dedekind domain. For a polynomial $f = \sum_{i=0}^{n} a_i T^i \in A[T]$, define its *content ideal* by $c(f) = (a_0, \dots, a_n)$. This is an ideal, whereas the content used in Gauss's lemma in Lecture 10 is a greatest common divisor of coefficients, an element determined up to a unit. Prove $c(fg) = c(f)c(g)$ for all $f, g \in A[T]$.

Exercise 10.11 (When a valuation ring is Noetherian; AM Chapter 9, Exercise 3)**.** Prove that a valuation ring which is not a field is Noetherian if and only if it is a DVR.

Exercise 10.12 (A principal maximal ideal with zero intersection of its powers; AM Chapter 9, Exercise 4)**.** Let $(A, \mathfrak{m})$ be a local domain, not a field. Suppose $\mathfrak{m} = (\pi)$ and $\bigcap_{n \geq 1} \mathfrak{m}^n = 0$. Prove that A is a DVR, without assuming it is Noetherian.

Exercise 10.13 (Flat modules over a Dedekind domain; AM Chapter 9, Exercise 5)**.** Let M be a finitely generated module over a Dedekind domain A. Prove that M is flat if and only if it is torsion-free. Here torsion-free means $am = 0$, $a \neq 0$, implies $m = 0$.

Exercise 10.14 (Two generators suffice; AM Chapter 9, Exercise 7)**.** For a nonzero ideal I in a Dedekind domain A, prove that every ideal of A/I is principal. Deduce that every ideal of A is generated by at most two elements.

Exercise 10.15 (Distributivity of ideal arithmetic; AM Chapter 9, Exercise 8)**.** For ideals I, J, K of a Dedekind domain, prove

$$I \cap (J + K) = (I \cap J) + (I \cap K), \qquad I + (J \cap K) = (I + J) \cap (I + K).$$

Exercise 10.16 (Compatible congruences without coprimality; AM Chapter 9, Exercise 9)**.** Let A be a Dedekind domain, $I_1, \ldots, I_n$ ideals of A, and $a_1, \ldots, a_n \in A$. Prove that the congruences $x \equiv a_i \pmod{I_i}$ have a simultaneous solution if and only if

$$a_i - a_j \in I_i + I_j \quad \text{for every } i, j.$$

Lecture 11
Graded rings and Hilbert functions

A polynomial is a sum of terms of different degrees. A graded ring records this decomposition algebraically, and a Hilbert function counts the independent terms of each degree. We first develop these constructions for polynomials and then for modules. The final sections concern powers of an ideal and their intersections with submodules; their proofs provide the algebraic tools for dimension theory in Lecture 12.

11.1 Graded rings and modules

In $k[x,y]$,

$$1+x+y+xy+x^3 = \underbrace{1}_{\deg 0} + \underbrace{(x+y)}_{\deg 1} + \underbrace{xy}_{\deg 2} + \underbrace{x^3}_{\deg 3}.$$

The decomposition by degree is unique, and multiplication adds degrees. A grading records these two properties.

Definition 11.1. A nonnegatively *graded ring* is a ring with an additive direct-sum decomposition

$$R = \bigoplus_{n\geq 0} R_n, \qquad R_i R_j \subseteq R_{i+j}, \qquad 1 \in R_0.$$

An element of R_n is *homogeneous of degree n*; the zero element belongs to every component. A *graded R-module* is an R-module with an additive decomposition $M = \bigoplus_{n\in\mathbb{Z}} M_n$ satisfying $R_i M_j \subseteq M_{i+j}$. The direct-sum condition means that each element has a unique expression as a sum of homogeneous components, with only finitely many nonzero components.

A homomorphism of graded modules is *degree-preserving* if it maps M_n into the degree-n component of its target for every n. A submodule $N \subseteq M$ is *graded* if it contains the homogeneous components of all its elements; equivalently, $N = \bigoplus_n (N \cap M_n)$. Then M/N has degree-n component $M_n/(N \cap M_n)$. Kernels and images of degree-preserving maps are graded: apply the map separately to the homogeneous components and use uniqueness of their decomposition.

The ring is *standard graded over* R_0 if it is generated as an R_0-algebra by finitely many elements of degree 1. In our first examples $R_0 = k$ is a field. The polynomial ring $S = k[x_1, \dots, x_r]$ with $\deg x_i = 1$ is the model. Its degree-n piece has basis the monomials $x_1^{a_1} \cdots x_r^{a_r}$ with $a_1 + \cdots + a_r = n$.

An ideal I is *homogeneous* if the homogeneous pieces of each of its elements also lie in I. Equivalently, it can be generated by homogeneous elements. One direction follows by decomposing the coefficients of a linear combination of homogeneous generators; the other follows because all the homogeneous elements of I already generate it. Only for such ideals does the quotient inherit the grading

$$(R/I)_n = R_n/(I \cap R_n).$$

For instance, $(x^2 - y^2)$ is homogeneous in the standard grading; $(x - 1)$ is not. In the second quotient, $x = 1$ would identify a degree-1 element with a degree-0 element. This contradicts a direct-sum decomposition unless the element is zero, which it is not.

Example 11.2 (A weighted grading)**.** The cusp equation $y^2 - x^3$ is not homogeneous when both variables have degree 1. It becomes homogeneous of degree 6 if $\deg x = 2$, $\deg y = 3$. The ring $k[t^2, t^3]$ then inherits its grading from $\deg t = 1$. The same ring can therefore be graded in a way adapted to its equation, but standard-graded results must not be applied without checking the degrees.

11.2 Hilbert functions and generating series

Let M be a finitely generated graded module over a standard graded k-algebra. We can choose finitely many homogeneous generators: start with any finite generating set and take all its homogeneous components. Each M_n is finite-dimensional. Indeed, only finitely many monomials in the algebra generators can multiply a fixed module generator into degree n. Moreover $M_n = 0$ for all sufficiently negative n.

Definition 11.3. The *Hilbert function* and *Hilbert series* of M are

$$h_M(n) = \dim_k M_n, \qquad H_M(z) = \sum_{n \in \mathbb{Z}} h_M(n) z^n.$$

Here z is an indeterminate. A *formal power series* is an expression $\sum_{n \geq 0} a_n z^n$, regarded as its sequence of coefficients. Two such series are equal when their coefficients agree. Addition is coefficientwise, and multiplication is defined by the finite sums

$$[z^n] \left(\sum_{i \geq 0} a_i z^i \right) \left(\sum_{j \geq 0} b_j z^j \right) = \sum_{i=0}^{n} a_i b_{n-i},$$

where $[z^n]$ denotes the coefficient of z^n. A *formal Laurent series* also permits finitely many negative powers. The same operations apply after multiplying by a sufficiently large power of z. Thus no numerical value of z, or convergence condition, is involved.

The identity $(1-z)(1+z+z^2+\cdots)=1$ is verified coefficient by coefficient. Accordingly $1/(1-z)$ denotes this formal inverse. For $S=k[x_1,\dots,x_r]$, choosing an exponent independently for each variable gives

$$H_S(z)=(1+z+z^2+\cdots)^r=\frac{1}{(1-z)^r}.$$

The coefficient is

$$h_S(n)=\binom{n+r-1}{r-1}\quad (n\geq 0,\ r\geq 1).$$

One proof counts nonnegative integer solutions to $a_1+\cdots+a_r=n$: arrange n marks and $r-1$ separators, with the numbers of marks between separators giving the exponents. Thus a ring in three variables has $\binom{n+2}{2}$ monomials of degree n.

If $0\to M'\to M\to M''\to 0$ is an exact sequence of graded modules with degree-preserving maps, it is exact in each degree. Additivity of vector-space dimensions gives

$$H_M(z)=H_{M'}(z)+H_{M''}(z).$$

Degree shifts make multiplication maps degree-preserving. Our convention is

$$M(-d)_n=M_{n-d},\qquad H_{M(-d)}(z)=z^dH_M(z).$$

In particular, the element 1 of $S(-d)$ has degree d.

Example 11.4 (One homogeneous equation)**.** Let $0\neq f\in S$ be homogeneous of degree $d>0$. Since S is a domain, multiplication by f is injective, and

$$0\longrightarrow S(-d)\xrightarrow{\cdot f} S\longrightarrow S/(f)\longrightarrow 0$$

is exact. Hence

$$H_{S/(f)}(z)=\frac{1-z^d}{(1-z)^r}.$$

For $S=k[x,y]$ and $f=xy$, this is $(1+z)/(1-z)=1+2z+2z^2+\cdots$, agreeing with the monomial count above.

Reading coefficients rather than only simplifying fractions

Consider $R=k[x,y]/(x^2)$ with both variables of degree one. For $n\geq 1$, the surviving degree-n monomials are y^n and xy^{n-1}, so

$$\dim_k R_0=1,\qquad \dim_k R_n=2\ (n\geq 1),\qquad H_R(t)=\frac{1+t}{1-t}.$$

Its Hilbert polynomial is the constant two. The cumulative count through degree n is instead $1+2n$. Summing graded pieces introduces another factor $(1-t)^{-1}$ in the generating series. Confusing the graded count with the cumulative count would give a dimension error of one.

For the shifted module $R(-a)$, the convention is $R(-a)_n = R_{n-a}$. Thus $H_{R(-a)}(t) = t^a H_R(t)$. In particular the injective map used for a homogeneous equation f of degree a is

$$R(-a) \xrightarrow{f} R,$$

which preserves degree: an element of original degree $n-a$ is sent to degree n. This explains both the sign of the shift and the numerator $1-t^a$ in hypersurface formulas.

The assumption of degree-one generators in eventual polynomiality is essential. If $k[u]$ is graded by $\deg u = 2$, its Hilbert function is one in even degrees and zero in odd degrees. Its series $1/(1-t^2)$ is rational, but its coefficients do not eventually agree with a polynomial. Indeed a polynomial vanishing at every sufficiently large odd integer is the zero polynomial, which cannot take value one at the even integers. Weighted gradings give periodic phenomena which the standard grading avoids.

11.3 The general Hilbert–Serre theorem

The coefficient ring in an associated graded construction need not be a field. We therefore also count homogeneous components by length. Recall from Lecture 9 that $\operatorname{length}_B(N)$ is the number of simple factors in a composition series of N; this number is independent of the series. Every finite module over an Artinian ring has finite length, and length is additive in short exact sequences.

More generally, an *additive function* λ on finite modules over a ring B assigns an integer to each isomorphism class and satisfies

$$\lambda(N) = \lambda(N') + \lambda(N'') \quad \text{if } 0 \to N' \to N \to N'' \to 0 \text{ is exact.}$$

In particular $\lambda(0) = 0$. Dimension over a field and length over an Artinian ring are the examples used below.

Proposition 11.5 (Graded Noetherian rings; AM 10.7)**.** *For a nonnegatively graded ring R, the following conditions are equivalent: R is Noetherian; and R_0 is Noetherian and R is generated by finitely many elements as an R_0-algebra.*

Proof. Suppose R is Noetherian. Choose homogeneous generators

$$R_+ = (x_1, \dots, x_s), \qquad \deg x_i = w_i > 0.$$

The quotient $R_0 \cong R/R_+$ is Noetherian. Put $B = R_0[x_1, \dots, x_s]$. Induct on n to show $R_n \subseteq B$. For $n = 0$ this is immediate. For $y \in R_n$, $n > 0$, write $y = \sum_i a_i x_i$ and take degree-n components:

$$y = \sum_i (a_i)_{n-w_i} x_i.$$

Each nonzero coefficient has degree $n - w_i < n$, hence lies in B by induction. Thus $y \in B$ and $R = B$.

Conversely, if R_0 is Noetherian and $R = R_0[x_1, \dots, x_s]$, evaluation gives

$$R_0[T_1, \dots, T_s] \twoheadrightarrow R.$$

Hilbert's basis theorem and the quotient criterion make R Noetherian. □

Fix a Noetherian graded ring $R = R_0[x_1, \dots, x_s]$, with $\deg x_i = w_i > 0$, and a finite nonnegatively graded R-module M. Each M_n is a finite R_0-module: homogeneous generators of M, multiplied by the finitely many monomials of each specified weighted degree, generate it. Put

$$P_\lambda(M,t) = \sum_{n \geq 0} \lambda(M_n) t^n.$$

Allowing finitely many negative degrees only changes the numerator below to a Laurent polynomial.

Theorem 11.6 (Hilbert–Serre; AM 11.1)**.** *There is $f(t) \in \mathbb{Z}[t]$ such that*

$$P_\lambda(M,t) = \frac{f(t)}{\prod_{i=1}^{s}(1 - t^{w_i})}.$$

Proof. Induct on s. For $s = 0$, all generators have degree zero. A finite homogeneous generating list for M therefore gives $M_n = 0$ outside finitely many degrees, so $P_\lambda(M,t)$ is a polynomial.

For $s > 0$, put $x = x_s$, $w = w_s$, and

$$K = (0 :_M x), \qquad C = M/xM.$$

Both are finite graded modules killed by x: K is finite because M is Noetherian, and C is a quotient of M. Thus both are finite over $R_0[x_1, \dots, x_{s-1}]$.

Degreewise exactness gives

$$0 \to K_{n-w} \to M_{n-w} \xrightarrow{x} M_n \to C_n \to 0.$$

Additivity and summation give

$$\begin{aligned} \lambda(M_n) - \lambda(M_{n-w}) &= \lambda(C_n) - \lambda(K_{n-w}), \\ (1 - t^w) P_\lambda(M,t) &= P_\lambda(C,t) - t^w P_\lambda(K,t). \end{aligned}$$

By induction, write $P_\lambda(C,t) = f_C(t)/D(t)$ and $P_\lambda(K,t) = f_K(t)/D(t)$, where

$$D(t) = \prod_{i=1}^{s-1}(1 - t^{w_i}).$$

Then

$$P_\lambda(M,t) = \frac{f_C(t) - t^{w_s} f_K(t)}{\prod_{i=1}^{s}(1 - t^{w_i})},$$

with integral polynomial numerator, as required. □

The order at $t = 1$ of a rational function is an algebraic notion. Factor all powers of $1-t$ from its numerator and denominator. A nonzero rational function then has the form

$$(1-t)^{-d}\frac{a(t)}{b(t)}, \qquad a(1)b(1) \neq 0,$$

for a unique integer d. We call d its *signed pole order*. When $d > 0$ the function has a pole of order d; when $d = 0$ it has a nonzero value at 1; when $d < 0$ it has a zero of order $-d$. For example, $(1+t)/(1-t)^2$ has pole order two. This definition involves only factorization in $\mathbb{Q}[t]$, without an analytic limit or a convergence assumption. The identically zero series is treated separately.

Corollary 11.7 (Eventual polynomiality; AM 11.2)**.** *If every $w_i = 1$, the function $\lambda(M_n)$ is eventually polynomial in n. If $P_\lambda(M,t)$ has a pole of order $d \geq 1$, this polynomial has degree exactly $d-1$. If the series has no pole, its coefficients are eventually zero.*

Proof. Cancel the common factors $1-t$. When a pole remains, write the result as $f(t)/(1-t)^d$ with $f(1) \neq 0$, and write $f(t) = \sum_{j=0}^{N} a_j t^j$. The geometric-series identity, differentiated or multiplied repeatedly, gives

$$(1-t)^{-d} = \sum_{n \geq 0} \binom{n+d-1}{d-1} t^n.$$

Consequently, for $n \geq N$,

$$\lambda(M_n) = \sum_{j=0}^{N} a_j \binom{n-j+d-1}{d-1}.$$

Each summand is a polynomial of degree at most $d-1$, and the leading coefficient of the sum is $f(1)/(d-1)! \neq 0$. If there is no pole, the reduced rational expression is a polynomial because its only possible denominator factors were $1-t$. Its coefficients eventually vanish. □

The eventual polynomial in Corollary 11.7 is the *Hilbert polynomial* for the chosen additive function; for a graded algebra over a field we take $\lambda = \dim_k$. It is unique, because two polynomials that agree at all sufficiently large integers have a difference with infinitely many roots. Its values at small integers need not equal the Hilbert function.

An element $x \in R$ is a *non-zero-divisor on M* if $xm = 0$ implies $m = 0$. This is precisely the assertion that multiplication by x on M is injective. It refers to the module M, not merely to the ring R.

Proposition 11.8 (Removing a homogeneous non-zero-divisor; AM 11.3)**.** *Let $x \in R_w$ with $w > 0$, and suppose multiplication by x on M is injective. Then*

$$P_\lambda(M/xM,t) = (1-t^w)P_\lambda(M,t).$$

Writing $d(N)$ for the signed pole order of $P_\lambda(N,t)$, we therefore have $d(M/xM) = d(M)-1$ whenever the series is nonzero. In particular this is the book's pole-order formula in its applications to length.

Proof. The exact sequence is now $0 \to M(-w) \xrightarrow{x} M \to M/xM \to 0$, so additivity proves the identity. The factor $1-t^w = (1-t)(1+t+\cdots+t^{w-1})$ has a simple zero at 1, since the second factor has value $w \neq 0$ as a rational number. Multiplication therefore reduces the signed pole order by exactly one. The positive-degree hypothesis matters; a degree-zero unit has zero quotient. □

Example 11.9 (Length over a coefficient ring)**.** If B is a nonzero Artinian ring and $R = B[x_1,\dots,x_s]$, with all variables of degree one, then

$$P_{\mathrm{length}_B}(R,t) = \frac{\mathrm{length}_B(B)}{(1-t)^s}.$$

Each monomial contributes one free copy of B, hence contributes $\mathrm{length}_B(B)$, not necessarily one. For example $B = k[\varepsilon]/(\varepsilon^2)$ has length two, so the numerator is 2. This normalization does not affect the pole order s.

11.4 Filtrations and associated graded rings

Definition 11.10. A *descending filtration* of an A-module M is a sequence of submodules $M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots$. For an ideal $I \subseteq A$, it is an *I-filtration* if $IM_n \subseteq M_{n+1}$ for every $n \geq 0$. Here IM_n consists of finite sums of products am with $a \in I$ and $m \in M_n$. The basic example is $M_n = I^n M$, where $I^0 = A$.

In particular the powers of I form an I-filtration of A. The *associated graded ring* of these powers is

$$\mathrm{gr}_I(A) = \bigoplus_{n\geq 0} I^n/I^{n+1}.$$

For $a \in I^i$ and $b \in I^j$, define the product of their classes to be the class of ab in I^{i+j}/I^{i+j+1}. If a is changed by $a' \in I^{i+1}$, the product changes by $a'b \in I^{i+j+1}$, so it is independent of the representatives. The same check handles b.

For an A-module M, the associated graded module for the powers of I is

$$\mathrm{gr}_I(M) = \bigoplus_{n\geq 0} I^n M/I^{n+1}M.$$

The action of $\mathrm{gr}_I(A)$ is given by $(a \bmod I^{i+1})(m \bmod I^{j+1}M) = am \bmod I^{i+j+1}M$. The same calculation with changed representatives proves that this is well-defined. More generally, for an I-filtration (M_n) we write

$$G(M) = \bigoplus_{n\geq 0} M_n/M_{n+1}.$$

Since $I^i M_j \subseteq M_{i+j}$, the identical rule makes $G(M)$ a graded module over $\operatorname{gr}_I(A)$. We also use the abbreviation $G_I(A) = \operatorname{gr}_I(A)$.

When $I = (a_1, \dots, a_r)$, the degree-1 classes of the a_i generate $\operatorname{gr}_I(A)$ as an algebra over A/I. Indeed, I^n is generated by products of n of the a_i, and coefficients can be reduced modulo I when working modulo I^{n+1}.

In a Noetherian local ring $(A, \mathfrak{m}, \kappa)$, this makes $\operatorname{gr}_{\mathfrak{m}}(A)$ a standard graded algebra over the residue field κ. For $A = k[t]_{(t)}$, its degree-n piece has basis the class of t^n, and

$$\operatorname{gr}_{(t)}(A) \cong k[T], \qquad \deg T = 1.$$

For $A = k[t]_{(t)}/(t^r)$, the corresponding ring is $k[T]/(T^r)$. The associated graded ring remembers successive orders of vanishing while discarding interactions between different orders. It is not in general the same ring as A.

11.5 Stable filtrations and the Artin–Rees lemma

An I-filtration (M_n) is *stable* if $IM_n = M_{n+1}$ for all sufficiently large n. Equivalently, there is an integer $c \geq 0$ such that $M_{c+n} = I^n M_c$ for all $n \geq 0$, by repeated substitution. The filtration $(I^n M)$ is stable from its first term. Two filtrations have *bounded difference* if, after shifting the index by a fixed integer, each is contained in the other as in the following lemma.

Lemma 11.11 (Stable filtrations differ by a bounded shift; AM 10.6)**.** *Two stable I-filtrations (M_n) and (M'_n) have bounded difference: there is $c \geq 0$ such that $M_{n+c} \subseteq M'_n$ and $M'_{n+c} \subseteq M_n$ for every $n \geq 0$.*

Proof. Choose $c, c' \geq 0$ such that

$$M_{n+c} = I^n M_c, \qquad M'_{n+c'} = I^n M'_{c'} \qquad (n \geq 0).$$

The filtration condition and stability give

$$M_{n+c} \subseteq I^n M \subseteq M'_n, \qquad M'_{n+c'} \subseteq I^n M \subseteq M_n.$$

Replacing c, c' by their maximum preserves these inclusions and gives the common bound. □

For a new indeterminate T, let $M[T] = \bigoplus_{n \geq 0} M T^n$ be the module of polynomials with coefficients in M; thus an element is a finite sum $\sum m_n T^n$. Polynomial multiplication makes this an $A[T]$-module. Define the *Rees ring* of I and the *Rees module* of the filtration by

$$\mathscr{R}_I(A) = \bigoplus_{n \geq 0} I^n T^n \subseteq A[T], \qquad \mathscr{R}(M_\bullet) = \bigoplus_{n \geq 0} M_n T^n.$$

The latter is the graded submodule $\bigoplus M_n T^n \subseteq M[T]$, with degree-$n$ coefficients restricted to M_n. The filtration condition implies $I^i M_j \subseteq M_{i+j}$, so multiplication makes it a module

over the Rees ring. If A is Noetherian and $I=(a_1,\dots,a_s)$, then $\mathscr{R}_I(A)=A[a_1T,\dots,a_sT]$, so it is Noetherian.

Lemma 11.12 (Stability is finite generation; AM 10.8)**.** *Let A be Noetherian, M finite, and (M_n) an I-filtration. Then $\mathscr{R}(M_\bullet)$ is finite over $\mathscr{R}_I(A)$ if and only if the filtration is stable.*

Proof. If $M_{c+n}=I^nM_c$, choose finite A-generating lists for $M_0,\dots,M_c$ and place them in their respective degrees. Then

$$\mathscr{R}(M_\bullet)=\sum_{j=0}^{c}\mathscr{R}_I(A)\,(M_jT^j).$$

For $n>c$, this equality uses $M_n=I^{n-c}M_c$; for $n\le c$, the degree-n list suffices. Hence the Rees module is finite.

Conversely, choose homogeneous generators $m_iT^{d_i}$ and put $c=\max_i d_i$. For $n\ge c$, the degree-n and degree-$(n+1)$ components give

$$M_n=\sum_i I^{n-d_i}m_i,$$
$$M_{n+1}=\sum_i I^{n+1-d_i}m_i=I\sum_i I^{n-d_i}m_i=IM_n.$$

Thus the filtration is stable. If the Rees module is zero, all M_n vanish and stability is immediate. □

Proposition 11.13 (Artin–Rees, stable-filtration form; AM 10.9)**.** *Let A be Noetherian, M finite, (M_n) a stable I-filtration, and $N\subseteq M$ a submodule. Then $(N\cap M_n)$ is a stable I-filtration of N.*

Proof. Put $N_n=N\cap M_n$. Then

$$IN_n\subseteq N\cap IM_n\subseteq N\cap M_{n+1}=N_{n+1},$$

so (N_n) is an I-filtration of N. Moreover

$$\mathscr{R}(N_\bullet)\subseteq\mathscr{R}(M_\bullet).$$

The module on the right is finite over the Noetherian ring $\mathscr{R}_I(A)$, hence its submodule on the left is finite. Since N is finite over A, Lemma 11.12 applies and proves stability. □

Corollary 11.14 (Artin–Rees, intersection formula; AM 10.10)**.** *Under the preceding finiteness assumptions there is $c\ge 0$ such that*

$$I^nM\cap N=I^{n-c}(I^cM\cap N)\qquad(n\ge c).$$

Proof. Apply Proposition 11.13 to $M_n=I^nM$. Choose c after which its induced filtration on N is stable, and iterate its equality $n-c$ times. □

Theorem 11.15 (The submodule filtration has bounded difference; AM 10.11)**.** *Let A be Noetherian, I an ideal, M a finite A-module, and $N \subseteq M$ a submodule. The two filtrations $(I^n N)$ and $(I^n M \cap N)$ have bounded difference. More precisely, there is $c \geq 0$ such that for every $n \geq 0$,*

$$I^{n+c}M \cap N \;\subseteq\; I^n N \;\subseteq\; I^n M \cap N.$$

Proof. The second inclusion follows from $N \subseteq M$. Choose the Artin–Rees bound c in Corollary 11.14. Its formula at index $n+c$ gives

$$I^{n+c}M \cap N = I^n(I^c M \cap N) \subseteq I^n N,$$

which is the first inclusion. Thus, for each fixed term of either filtration, every sufficiently late term of the other is contained in it. The uniform integer c gives the stated quantitative comparison. □

Theorem 11.16 (Krull's intersection theorem; AM 10.17)**.** *Let M be finite over a Noetherian ring. For an ideal $I \subseteq A$, one has*

$$E = \bigcap_{n \geq 0} I^n M = \{m \in M : (1-a)m = 0 \text{ for some } a \in I\}.$$

Indeed a single element $1-a$ annihilates the whole module E.

Proof. The submodule E is finite, with $E \cap I^n M = E$ for every n. Artin–Rees makes this constant filtration stable; hence $IE = E$. Write

$$E = Ae_1 + \cdots + Ae_r, \qquad e_i = \sum_j c_{ij} e_j, \qquad c_{ij} \in I.$$

For $C = (c_{ij})$ and $\mathbf{e} = (e_i)^{\mathsf{t}}$, the determinant argument gives

$$(1-C)\mathbf{e} = 0 \quad \Longrightarrow \quad \det(1-C)\mathbf{e} = 0.$$

Since $\det(1-C) \equiv 1 \pmod{I}$, write it as $1-a$ with $a \in I$. Then $(1-a)E = 0$. If $E = 0$, take $a = 0$.

Conversely,

$$(1-a)m = 0, \quad a \in I \quad \Longrightarrow \quad m = am = a^2 m = \cdots = a^n m \in I^n M$$

for every $n \geq 0$. Thus $m \in E$, proving both assertions. □

Corollary 11.17 (Intersection in a domain; AM 10.18)**.** *If A is a Noetherian domain and $I \neq A$, then $\bigcap_n I^n = 0$.*

Proof. No element $1-a$, $a \in I$, is zero, since otherwise $1 \in I$. Such an element cannot kill a nonzero element of a domain. Apply Theorem 11.16 with $M = A$. □

Corollary 11.18 (Intersection for a Jacobson-radical ideal; AM 10.19)**.** *If I lies in the Jacobson radical of a Noetherian ring and M is finite, then $\bigcap_n I^n M = 0$.*

Proof. Every $1-a$ with $a \in I$ is a unit. The intersection described in Theorem 11.16 is therefore zero. □

Corollary 11.19 (Intersection in a local ring; AM 10.20)**.** *For a finite module M over a Noetherian local ring $(A,\mathfrak{m})$, one has $\bigcap_{n\geq 0}\mathfrak{m}^n M = 0$. In particular $\bigcap_{n\geq 0}\mathfrak{m}^n = 0$.*

Proof. The Jacobson radical of a local ring is its unique maximal ideal. Apply Corollary 11.18, first to M and then to A. □

Corollary 11.20 (Primary ideals and the localization kernel; AM 10.21)**.** *Let A be Noetherian and $\mathfrak{p}$ a prime ideal. The intersection of all $\mathfrak{p}$-primary ideals of A is the kernel of the localization map $A \to A_{\mathfrak{p}}$.*

Proof. Recall that Q is $\mathfrak{p}$-primary when it is primary and $\sqrt{Q} = \mathfrak{p}$. If x belongs to the localization kernel, then $sx = 0$ for some $s \notin \mathfrak{p}$. For any $\mathfrak{p}$-primary Q, we have $sx \in Q$. If $x \notin Q$, primaryness would force a power of s into Q, contradicting $s \notin \sqrt{Q}$. Thus the kernel lies in every such Q.

For the reverse inclusion, put

$$Q_n = \{a \in A : a/1 \in (\mathfrak{p}A_{\mathfrak{p}})^n\}, \qquad n \geq 1.$$

The ideal $(\mathfrak{p}A_{\mathfrak{p}})^n$ has maximal radical in $A_{\mathfrak{p}}$ and is primary by the elementary criterion in Proposition 9.37. Contracting a primary ideal preserves its defining implication, and taking radicals commutes with contraction. Hence each Q_n is $\mathfrak{p}$-primary. An element in every $\mathfrak{p}$-primary ideal belongs in particular to every Q_n. Its image in $A_{\mathfrak{p}}$ lies in $\bigcap_n (\mathfrak{p}A_{\mathfrak{p}})^n = 0$ by Corollary 11.19. It therefore belongs to the localization kernel. The two inclusions prove equality without a decomposition theorem. □

Example 11.21 (The shift in Artin–Rees can be visible)**.** Take $A = k[x]$, $I = (x)$, $M = A$, and $N = (x^r)$. Then

$$I^n N = (x^{n+r}), \qquad I^n M \cap N = (x^{\max(n,r)}).$$

The filtrations differ at small indices. For $n \geq r$, however, $I^n M \cap N = I^{n-r}(I^r M \cap N)$, exactly as the theorem predicts with $c = r$. The theorem does not say that intersecting commutes with multiplying by I at every index.

Lemma 11.22 (Finite associated graded modules)**.** *Let A be Noetherian, I an ideal, and M a finite module with a stable I-filtration (M_n). Then $G_I(A) = \bigoplus I^n/I^{n+1}$ is a Noetherian graded ring, and $G(M) = \bigoplus M_n/M_{n+1}$ is finite over it.*

Proof. The degree-one classes of any finite list generating I generate $G_I(A)$ over A/I. Thus $G_I(A)$ is a finite type algebra over a Noetherian ring and is Noetherian. Choose c with $M_{c+n} = I^n M_c$ for all $n \geq 0$. Each of the finitely many pieces M_j/M_{j+1}, $0 \leq j \leq c$, is finite over A/I: it is a quotient of a finite A-module and is killed by I. Choose generators for all these pieces. Every piece of degree greater than c is generated from degree c, using the displayed stability equality. Hence the combined finite list generates $G(M)$. □

The lemma is the algebraic finiteness assertion used in the proof of AM 10.22. Its proof here uses only finite generation and stability.

Why submodules need the induced filtration

Even an inclusion need not remain injective on associated graded modules if each module is given its own powers filtration. Let $A = k[x]$, $I = (x)$, $M = A$, and $N = xA \subseteq M$. Then

$$I^n N = x^{n+1}A, \qquad \operatorname{gr}_I(N)_n = x^{n+1}A/x^{n+2}A.$$

The map to $\operatorname{gr}_I(M)_n = x^nA/x^{n+1}A$ induced by inclusion is zero in every degree, although each source piece is one-dimensional over k. The original inclusion was injective; the choice of filtration has changed what its leading term means.

For an exact sequence $0 \to N \to M \to Q \to 0$, start instead with a filtration M_n, and use

$$N_n = N \cap M_n, \qquad Q_n = (M_n + N)/N.$$

Then in every degree n the sequence

$$0 \to N_n/N_{n+1} \to M_n/M_{n+1} \to Q_n/Q_{n+1} \to 0$$

is exact. For the kernel assertion, if $m \in M_n$ maps to Q_{n+1}, write $m = m' + u$ with $m' \in M_{n+1}$ and $u \in N$. Since both m, m' are in M_n, the difference u lies in N_n and represents the same graded class. The remaining assertions follow directly from the definitions. Artin–Rees ensures that for finite modules over a Noetherian ring, the induced filtration and the powers filtration differ by only a bounded shift. They need not coincide term by term.

11.6 Exercises

Exercise 11.1. Give bases for the degree-n pieces of $R = k[x,y]/(xy)$ with the standard grading, and determine their dimensions.

Exercise 11.2. Compute the Hilbert series of $k[x,y]/(x^2, y^3)$ and its eventual Hilbert polynomial.

Exercise 11.3. Give $k[u]$ the grading $\deg u = 2$. Find its Hilbert series and show that its Hilbert function is not eventually a polynomial in n.

Exercise 11.4. Let $A = k[\varepsilon]/(\varepsilon^3)$ and $\mathfrak{m} = (\varepsilon)$. Compute $\operatorname{gr}_{\mathfrak{m}}(A)$, and explain why its degree-0 piece is k rather than A.

Exercise 11.5 (Why the Jacobson-radical hypothesis matters)**.** For $A = k \times k$ and $I = k \times 0$, compute $N = \bigcap_{n \geq 1} I^n$ and find an element $a \in I$ such that $(1-a)N = 0$, as asserted by Krull's intersection theorem.

Lecture 12
Dimension theory and regular local rings

A line has one independent coordinate and a plane has two. The algebraic definition of dimension will measure this distinction by chains of prime ideals. For a Noetherian local ring, we shall prove that this dimension also measures the growth of certain finite-length quotients and the number of equations needed to obtain such a quotient. The proofs use the graded constructions of Lecture 11. The main reference is [1, Chapter 11].

12.1 Prime chains and height

Definition 12.1. For a nonzero ring A, its *Krull dimension* is the supremum of the lengths of chains of prime ideals

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_d.$$

The length is d, the number of strict inclusions. The *height* $\operatorname{ht}(\mathfrak{p})$ is the supremum of lengths of chains ending at $\mathfrak{p}$. Neither supremum is assumed finite initially.

A field has dimension zero. The integers have dimension one: their primes are (0) and (p), with p an ordinary prime number. Likewise $k[t]$ has dimension one, because every nonzero prime ideal in this principal ideal domain is maximal. A nonzero Artinian ring has dimension zero, including rings such as $k[\varepsilon]/(\varepsilon^3)$ with nilpotents.

Proposition 12.2. *For any prime $\mathfrak{p}$ of a ring A,* $\dim A_{\mathfrak{p}} = \operatorname{ht}(\mathfrak{p})$. *Also*

$$\dim A = \sup_{\mathfrak{m}\ \textit{maximal}} \dim A_{\mathfrak{m}}, \qquad \dim A = \dim A/\sqrt{(0)}.$$

Proof. The localization correspondence is an inclusion-preserving bijection

$$\operatorname{Spec} A_{\mathfrak{p}} \longleftrightarrow \{\mathfrak{q} \in \operatorname{Spec} A : \mathfrak{q} \subseteq \mathfrak{p}\}.$$

It preserves strict inclusions. Every chain on the right extends to one ending at $\mathfrak{p}$, so $\dim A_{\mathfrak{p}} = \operatorname{ht}(\mathfrak{p})$.

Every prime is contained in a maximal ideal. Thus every finite prime chain extends to a chain ending at some maximal $\mathfrak{m}$, and

$$\dim A = \sup_{\mathfrak{m}\text{ maximal}} \operatorname{ht}(\mathfrak{m}) = \sup_{\mathfrak{m}\text{ maximal}} \dim A_{\mathfrak{m}}.$$

Finally every prime contains $\sqrt{0}$. The inclusion-preserving bijection

$$\operatorname{Spec}(A/\sqrt{0}) \longleftrightarrow \operatorname{Spec} A, \qquad \mathfrak{p}/\sqrt{0} \longleftrightarrow \mathfrak{p},$$

proves that quotienting by the nilradical leaves dimension unchanged. □

The local ring of $k[t]$ at (0) is the field $k(t)$ and has dimension zero; its local ring at $(t-a)$ has dimension one. Thus local dimension at an arbitrary prime need not equal global dimension.

12.2 The Hilbert–Samuel polynomial

Throughout this section $(A,\mathfrak{m})$ is a nonzero Noetherian local ring. An ideal $\mathfrak{q}$ with $\sqrt{\mathfrak{q}} = \mathfrak{m}$ is called $\mathfrak{m}$-*primary*. We use only this elementary maximal-radical condition. It says equivalently that

$$\mathfrak{m}^r \subseteq \mathfrak{q} \subseteq \mathfrak{m} \quad \text{for some } r \geq 1.$$

To see the nontrivial direction, choose generators $u_1,\dots,u_e$ of $\mathfrak{m}$ and exponents a_i with $u_i^{a_i} \in \mathfrak{q}$. Every product of $r = 1 + \sum(a_i - 1)$ generators has some u_i as a factor at least a_i times, so lies in $\mathfrak{q}$. The other direction follows by taking radicals.

The successive factors $\mathfrak{m}^i/\mathfrak{m}^{i+1}$ are finite-dimensional over $\kappa = A/\mathfrak{m}$. Since $A/\mathfrak{q}$ is a quotient of $A/\mathfrak{m}^r$,

$$\operatorname{length}_A(A/\mathfrak{q}) \leq \operatorname{length}_A(A/\mathfrak{m}^r) = \sum_{i=0}^{r-1} \dim_\kappa(\mathfrak{m}^i/\mathfrak{m}^{i+1}) < \infty.$$

Every simple factor is κ. If a subfield $k \subseteq A$ maps isomorphically onto κ, length consequently agrees with k-dimension. Length also applies when no such subfield exists, as for $\mathbb{Z}_{(p)}$ with residue field $\mathbf{F}_p$.

Proposition 12.3 (Hilbert–Samuel polynomial for a filtered module; AM 11.4)**.** *Let M be a finite A-module with a stable $\mathfrak{q}$-filtration (M_n), and let s be the least number of generators of $\mathfrak{q}$.*

1. *M/M_n has finite length for every n.*
2. *For all sufficiently large n, its length is a polynomial $g(n)$ of degree at most s.*
3. *For nonzero M, the degree and leading coefficient of this polynomial depend only on M and $\mathfrak{q}$, not on the chosen stable filtration. For $M = 0$ the polynomial is zero.*

Proof. By Lemma 11.22, $G(M) = \bigoplus_{j\geq 0} M_j/M_{j+1}$ is finite over $G_{\mathfrak{q}}(A)$. Its degree-zero coefficient ring $A/\mathfrak{q}$ is Artinian, so every graded piece has finite length. Additivity along

the finite filtration gives

$$\ell_n := \operatorname{length}_A(M/M_n) = \sum_{j=0}^{n-1} \operatorname{length}_A(M_j/M_{j+1}) < \infty.$$

This proves (i).

The classes of s generators of $\mathfrak{q}$ generate $G_{\mathfrak{q}}(A)$ in degree one. Hilbert–Serre therefore gives

$$\sum_{j\geq 0} \operatorname{length}_A(M_j/M_{j+1})t^j = \frac{f(t)}{(1-t)^s}, \qquad f(t) \in \mathbb{Z}[t].$$

Taking partial sums yields

$$\sum_{n\geq 0} \ell_n t^n = \frac{t}{1-t} \sum_{j\geq 0} \operatorname{length}_A(M_j/M_{j+1})t^j = \frac{tf(t)}{(1-t)^{s+1}}.$$

Corollary 11.7 makes ℓ_n eventually polynomial of degree at most s. For $s = 0$, the same formula gives eventual constancy.

For (iii), let (M'_n) be another stable filtration, giving a polynomial $g'(n)$. Lemma 11.11 supplies c such that

$$M_{n+c} \subseteq M'_n, \qquad M'_{n+c} \subseteq M_n.$$

Taking quotient lengths gives, for all large n,

$$g(n+c) \geq g'(n), \qquad g'(n+c) \geq g(n).$$

If $M \neq 0$, Nakayama gives $M/\mathfrak{q}M \neq 0$. Comparison with the $\mathfrak{q}$-adic filtration then makes both polynomials eventually positive. Write

$$g(n) = an^d + \text{lower-degree terms}, \qquad g'(n) = bn^e + \text{lower-degree terms}, \qquad a, b > 0.$$

The first inequality gives $d \geq e$ and the second $e \geq d$. Thus $d = e$. Dividing the two inequalities by n^d and letting n increase gives $a \geq b$ and $b \geq a$, hence $a = b$. For $M = 0$, all lengths are zero. □

Write $\chi_{\mathfrak{q}}^M(n)$ for the eventual polynomial of $\operatorname{length}(M/\mathfrak{q}^n M)$ and $\chi_{\mathfrak{q}}(n)$ when $M = A$. These are polynomials recording eventual values, not necessarily correct for every small n.

Corollary 12.4 (The characteristic polynomial; AM 11.5)**.** *For all large n,* $\operatorname{length}(A/\mathfrak{q}^n) = \chi_{\mathfrak{q}}(n)$ *is a polynomial of degree at most the least number of generators of* $\mathfrak{q}$.

Proof. Apply Proposition 12.3 to $M = A$ and $M_n = \mathfrak{q}^n$. □

Proposition 12.5 (The degree is independent of the ideal; AM 11.6)**.** *One has* $\deg \chi_{\mathfrak{q}} = \deg \chi_{\mathfrak{m}}$ *for every* $\mathfrak{m}$*-primary ideal* $\mathfrak{q}$.

Proof. Choose r with $\mathfrak{m}^r \subseteq \mathfrak{q} \subseteq \mathfrak{m}$. Then $\mathfrak{m}^{rn} \subseteq \mathfrak{q}^n \subseteq \mathfrak{m}^n$, so surjectivity of quotient maps gives

$$\chi_{\mathfrak{m}}(n) \leq \chi_{\mathfrak{q}}(n) \leq \chi_{\mathfrak{m}}(rn)$$

for all large n. These positive polynomials must have the same degree: replacing n by rn changes a leading coefficient but not a degree. □

Let $d(A) = \deg \chi_{\mathfrak{m}}$. This nonnegative integer is our growth measurement. Let $\delta(A)$ be the least number of generators of an $\mathfrak{m}$-primary ideal. This is finite since $\mathfrak{m}$ itself is such an ideal. An ideal generated by the empty list means the zero ideal; thus $\delta(A) = 0$ is permitted.

Proposition 12.6 (Equations bound growth; AM 11.7)**.** $\delta(A) \geq d(A)$.

Proof. Choose an $\mathfrak{m}$-primary ideal with $\delta(A)$ generators. Corollary 12.4 bounds its polynomial degree by $\delta(A)$, and Proposition 12.5 identifies that degree with $d(A)$. □

12.3 Growth under quotient by a non-zero-divisor

Proposition 12.7 (A quotient lowers the degree; AM 11.8)**.** *Let $M \neq 0$ be finite, and let x be a non-zero-divisor on M. If $M' = M/xM \neq 0$, then $\deg \chi_{\mathfrak{q}}^{M'} \leq \deg \chi_{\mathfrak{q}}^{M} - 1$. If $M' = 0$, its characteristic polynomial is zero and the assertion is interpreted as vanishing rather than assigning a degree to zero.*

Proof. Put $N = xM$. Injectivity of multiplication by x gives $N \cong M$. Give N the stable filtration $N_n = N \cap \mathfrak{q}^n M$, using Artin–Rees. At level n,

$$0 \longrightarrow N/N_n \longrightarrow M/\mathfrak{q}^n M \longrightarrow M'/\mathfrak{q}^n M' \longrightarrow 0$$

is exact: the kernel on the right is

$$(N + \mathfrak{q}^n M)/\mathfrak{q}^n M \cong N/(N \cap \mathfrak{q}^n M) = N/N_n.$$

Let $g(n)$ be the eventual length of N/N_n. Additivity gives

$$\chi_{\mathfrak{q}}^{M'} = \chi_{\mathfrak{q}}^{M} - g.$$

By Proposition 12.3(iii) and $N \cong M$, the two polynomials on the right have the same degree and leading coefficient. Their leading terms cancel. Thus the difference is zero or has degree at most $\deg \chi_{\mathfrak{q}}^{M} - 1$. □

Corollary 12.8 (Growth of a quotient by one element; AM 11.9)**.** *If x is a non-zero-divisor in A and is not a unit, then $d(A/(x)) \leq d(A) - 1$. If x is a unit, the quotient is the zero ring and there is no nonzero characteristic polynomial to compare.*

Proof. Apply the preceding proposition to $M = A$ and $\mathfrak{q} = \mathfrak{m}$. The quotient is nonzero because $x \in \mathfrak{m}$. Its lengths computed as an A-module are the same as its lengths over $A/(x)$: the composition factors on both sides are $A/\mathfrak{m}$. Its quotient filtration is given by powers of $\mathfrak{m}/(x)$, so its polynomial degree is $d(A/(x))$. □

Proposition 12.9 (Growth bounds every prime chain; AM 11.10)**.** $d(A) \geq \dim A$.

Proof. Induct on $d = d(A)$, simultaneously for all nonzero Noetherian local rings. If $d = 0$, the lengths $\operatorname{length}(A/\mathfrak{m}^n)$ are eventually constant. Exactness of

$$0 \to \mathfrak{m}^n/\mathfrak{m}^{n+1} \to A/\mathfrak{m}^{n+1} \to A/\mathfrak{m}^n \to 0$$

gives $\mathfrak{m}^n/\mathfrak{m}^{n+1} = 0$ for large n. Nakayama then gives

$$\mathfrak{m}^n = \mathfrak{m}\mathfrak{m}^n \quad \Longrightarrow \quad \mathfrak{m}^n = 0.$$

Every prime contains $\sqrt{\mathfrak{m}^n} = \mathfrak{m}$, so $\operatorname{Spec} A = \{\mathfrak{m}\}$ and $\dim A = 0$.

For $d > 0$, take a strict chain $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_r$. The assertion is immediate for $r = 0$. Otherwise choose $x \in \mathfrak{p}_1 \setminus \mathfrak{p}_0$ and put $B = A/\mathfrak{p}_0$. Its image $\bar{x}$ is a nonzero nonunit in the local domain B. Quotient maps and Corollary 12.8 give

$$\begin{aligned} \operatorname{length}\big(B/(\mathfrak{m}B)^n\big) &\le \operatorname{length}(A/\mathfrak{m}^n), \\ d(B) &\le d(A) = d, \\ d\big(B/(\bar{x})\big) &\le d(B) - 1 \le d - 1. \end{aligned}$$

Induction therefore gives $\dim B/(\bar{x}) \le d-1$. The images of $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ form a strict chain of length $r-1$ in this quotient. Hence

$$r - 1 \le \dim B/(\bar{x}) \le d - 1 \quad \Longrightarrow \quad r \le d.$$

Taking the supremum over prime chains proves $\dim A \le d(A)$. □

Corollary 12.10 (Local dimension is finite; AM 11.11)**.** *Every Noetherian local ring has finite Krull dimension.*

Proof. The integer $d(A)$ is finite and bounds the dimension by Proposition 12.9. □

Corollary 12.11 (Finite heights and descending prime chains; AM 11.12)**.** *In a Noetherian ring every prime has finite height. The set of prime ideals satisfies the descending chain condition.*

Proof. The localization $A_{\mathfrak{p}}$ is Noetherian local, so $\operatorname{ht}(\mathfrak{p}) = \dim A_{\mathfrak{p}} < \infty$. If a descending prime chain starting at $\mathfrak{p}$ failed to stabilize, deleting repetitions would yield arbitrarily long finite strict chains below $\mathfrak{p}$, contradicting its finite height. Notice that there need not be a uniform finite bound on the heights of all primes in a nonlocal Noetherian ring. □

12.4 Systems of parameters

Definition 12.12. Let $(A, \mathfrak{m})$ be a Noetherian local ring of dimension d, which is finite by Corollary 12.10. A *system of parameters* is a list $x_1, \ldots, x_d \in \mathfrak{m}$ such that $\sqrt{(x_1, \ldots, x_d)} = \mathfrak{m}$. Equivalently, the quotient by their ideal has finite length. This does not require the list to generate $\mathfrak{m}$. When $d = 0$ the list is empty and its generated ideal is (0).

Proposition 12.13 (Enough equations can always be chosen; AM 11.13)**.** *If A has dimension d, there is an $\mathfrak{m}$-primary ideal generated by d elements. Hence* $\dim A \geq \delta(A)$.

Proof. We inductively choose $x_1, \ldots, x_d \in \mathfrak{m}$ so that every prime containing $(x_1, \ldots, x_i)$ has height at least i. At $i = 0$ this holds without choosing anything. Suppose the first $i-1$ elements have been chosen. By Lemma 9.38, there are only finitely many primes minimal over their ideal. Among these, list those of height exactly $i-1$ as $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$.

Since $i - 1 < d = \operatorname{ht}(\mathfrak{m})$, none of these primes is $\mathfrak{m}$. Prime avoidance gives an element $x_i \in \mathfrak{m} \setminus \bigcup_j \mathfrak{p}_j$. If the list is empty, any element of $\mathfrak{m}$ will do. Now take a prime $\mathfrak{q}$ containing $(x_1, \ldots, x_i)$. It contains a prime $\mathfrak{p}$ minimal over $(x_1, \ldots, x_{i-1})$. The induction hypothesis gives $\operatorname{ht}(\mathfrak{p}) \geq i - 1$. If equality holds, $\mathfrak{p}$ is on our finite list, and $x_i \notin \mathfrak{p}$ but $x_i \in \mathfrak{q}$, so $\mathfrak{p} \subsetneq \mathfrak{q}$. Appending this strict inclusion to a chain of length $i-1$ gives $\operatorname{ht}(\mathfrak{q}) \geq i$. If $\operatorname{ht}(\mathfrak{p}) > i - 1$, the same bound follows immediately from $\mathfrak{p} \subseteq \mathfrak{q}$.

After d steps, every prime over $(x_1, \ldots, x_d)$ has height at least d. Such a prime must equal $\mathfrak{m}$: a strict inclusion into $\mathfrak{m}$ would create a chain of length $d+1$. Thus the radical of this ideal is $\mathfrak{m}$, as required. When $d = 0$, every prime is already $\mathfrak{m}$, so the zero ideal has radical $\mathfrak{m}$ and the empty list works. □

Theorem 12.14 (The local dimension theorem; AM 11.14)**.** *For a nonzero Noetherian local ring $(A, \mathfrak{m})$, the following three integers agree:*

1. $\dim A$*, the maximum length of strict prime chains;*
2. $\deg \chi_{\mathfrak{m}}$*, where* $\chi_{\mathfrak{m}}(n) = \operatorname{length}(A/\mathfrak{m}^n)$ *for all large n;*
3. *the least number of generators of an ideal with radical* $\mathfrak{m}$.

Proof. The preceding propositions give all three inequalities

$$\delta(A) \;\geq\; d(A) \;\geq\; \dim A \;\geq\; \delta(A).$$

They force equality. All quantities are finite, so the finite supremum of prime-chain lengths is attained. □

Proposition 12.13 proves that systems of parameters exist. The dimension theorem proves that no list with fewer elements can generate an ideal with radical $\mathfrak{m}$.

Example 12.15 (A line, a plane, and affine space)**.** For $A = k[t]_{(t)}$, the quotient $A/(t)^n$ has basis $1, t, \ldots, t^{n-1}$, so its length is n and $\dim A = 1$. A denominator nonzero at zero is invertible modulo t^n by a finite geometric-series calculation, which explains why localization adds no new basis elements.

For $A = k[x_1, \ldots, x_r]_{(x_1, \ldots, x_r)}$, the quotient by $\mathfrak{m}^n$ has basis the monomials of total degree less than n. Its length is

$$\binom{n+r-1}{r},$$

a polynomial of degree r. Thus this local ring has dimension r. Its associated graded ring is $k[X_1, \ldots, X_r]$: a homogeneous polynomial of degree j cannot acquire a term of smaller degree on multiplying by a denominator with nonzero constant term, so its degree-j monomial classes are independent as well as spanning.

12.5 Height bounds and the effect of one equation

Corollary 12.16 (Dimension is bounded by first-order directions; AM 11.15)**.** *For* $\kappa = A/\mathfrak{m}$*, one has* $\dim A \le \dim_\kappa(\mathfrak{m}/\mathfrak{m}^2)$.

Proof. Let $e = \dim_\kappa(\mathfrak{m}/\mathfrak{m}^2)$ and lift a basis to $x_1, \dots, x_e \in \mathfrak{m}$. For $N = \mathfrak{m}/(x_1, \dots, x_e)$,

$$N/\mathfrak{m}N \cong \frac{\mathfrak{m}}{\mathfrak{m}^2 + (x_1, \dots, x_e)} = 0.$$

Nakayama gives $N = 0$, hence $\mathfrak{m} = (x_1, \dots, x_e)$. Since $\sqrt{\mathfrak{m}} = \mathfrak{m}$, Theorem 12.14 yields $\dim A = \delta(A) \le e$. □

Corollary 12.17 (The generalized height theorem; AM 11.16)**.** *Let A be Noetherian, and let $\mathfrak{p}$ be a prime minimal over an ideal $(x_1, \dots, x_r)$. Then* $\operatorname{ht}(\mathfrak{p}) \le r$.

Proof. Put $I = (x_1, \dots, x_r)$. If a prime of $A_\mathfrak{p}$ contains $IA_\mathfrak{p}$, its contraction $\mathfrak{q}$ satisfies

$$I \subseteq \mathfrak{q} \subseteq \mathfrak{p}.$$

Minimality of $\mathfrak{p}$ forces $\mathfrak{q} = \mathfrak{p}$. Therefore

$$\sqrt{IA_\mathfrak{p}} = \mathfrak{p}A_\mathfrak{p}, \qquad IA_\mathfrak{p} = (x_1/1, \dots, x_r/1).$$

The dimension theorem in $A_\mathfrak{p}$ now gives

$$\operatorname{ht}(\mathfrak{p}) = \dim A_\mathfrak{p} = \delta(A_\mathfrak{p}) \le r.$$

□

Corollary 12.18 (Krull's principal ideal theorem; AM 11.17)**.** *If A is Noetherian and x is neither a unit nor a zero-divisor, every prime minimal over (x) has height exactly one.*

Proof. The preceding corollary bounds the height by one. Suppose such a prime $\mathfrak{p}$ had height zero. Then $A_\mathfrak{p}$ is Noetherian local of dimension zero. The dimension theorem, or the zero-dimensional Artinian criterion, says that its maximal ideal is nilpotent. Since $x/1 \in \mathfrak{p}A_\mathfrak{p}$, we obtain $(x/1)^n = 0$ for some n. By the equality criterion for localization, there is $s \notin \mathfrak{p}$ with $sx^n = 0$ in A. But multiplication by x^n is injective, being an iteration of multiplication by the non-zero-divisor x. Thus $s = 0$, contradicting $s \notin \mathfrak{p}$. The height cannot be zero and must be one. □

Corollary 12.19 (A non-zero-divisor cuts dimension by one; AM 11.18)**.** *For a non-zero-divisor $x \in \mathfrak{m}$,* $\dim A/(x) = \dim A - 1$.

Proof. Corollary 12.8 and the dimension theorem give

$$\dim A/(x) = d(A/(x)) \le d(A) - 1 = \dim A - 1.$$

Conversely, set $r = \dim A/(x)$ and choose a system of parameters $\bar{y}_1, \ldots, \bar{y}_r$ in $A/(x)$. Lifting to $y_i \in \mathfrak{m}$ gives

$$\sqrt{(x, y_1, \ldots, y_r)} = \mathfrak{m}.$$

This ideal has $r+1$ generators, so

$$\dim A = \delta(A) \leq r+1 = \dim A/(x) + 1.$$

The two inequalities prove equality. □

Example 12.20 (An equation can fail to lower dimension)**.** In $A = (k[x,y]/(xy))_{(x,y)}$, the element x is a zero-divisor because $xy = 0$ and $y \neq 0$. Every prime containing xy contains x or y. Hence a chain of such primes is contained in the primes above one of these two ideals and has length at most one; the chain $(x) \subsetneq (x,y)$ has length one. Thus $\dim A = 1$. Its quotient by x is $k[y]_{(y)}$, also of dimension one.

In contrast, the map

$$k[x,y]/(xy) \longrightarrow k[x] \times k[y], \qquad f \longmapsto (f(x,0), f(0,y))$$

is injective: its kernel before quotienting consists of polynomials divisible by both variables, hence by xy. It sends $x+y$ to (x,y). Multiplication by this pair is injective because the two factors are domains. Thus $x+y$ is a non-zero-divisor, also after localization. Substitution $y = -x$ identifies its quotient with $k[x]_{(x)}/(x^2)$, of dimension zero. These two calculations exhibit the role of the non-zero-divisor hypothesis.

12.6 Independence of parameters

A system of parameters satisfies a useful independence property: a homogeneous relation among its elements cannot have a unit coefficient. We first prove a polynomial lemma over an Artinian coefficient ring.

Lemma 12.21 (A unit coefficient over an Artinian local ring)**.** *Let $(B, \mathfrak{n})$ be an Artinian local ring. A polynomial $f \in B[t_1, \ldots, t_d]$ with at least one unit coefficient is a non-zero-divisor.*

Proof. The maximal ideal $\mathfrak{n}$ is nilpotent. Suppose g is a nonzero polynomial. Among the finitely many nonzero coefficients of g, choose the smallest order in the powers of $\mathfrak{n}$. More precisely, there is r such that every coefficient is in $\mathfrak{n}^r$ and at least one is not in $\mathfrak{n}^{r+1}$; existence follows from nilpotence. The class of g in $(\mathfrak{n}^r/\mathfrak{n}^{r+1})[t_1, \ldots, t_d]$ is therefore nonzero. Here $W[t_1, \ldots, t_d]$ for a module W denotes finite sums of monomials with coefficients in W, with the coefficientwise module operations. Thus the displayed expression is a polynomial module over the residue field $\kappa = B/\mathfrak{n}$. Choose a κ-basis of $\mathfrak{n}^r/\mathfrak{n}^{r+1}$, so the class of g becomes a tuple of ordinary polynomials, at least one nonzero. Reduction of f modulo $\mathfrak{n}$ is nonzero because it has a unit coefficient. Multiplication by this nonzero polynomial is injective on every coordinate, since $\kappa[t_1, \ldots, t_d]$ is a domain. Thus the class of fg is nonzero, and in particular $fg \neq 0$. □

Proposition 12.22 (Independence of parameters; AM 11.20)**.** *Let $x_1,\dots,x_d$ be a system of parameters in A, and put $\mathfrak{q}=(x_1,\dots,x_d)$. Let $f(t_1,\dots,t_d)$ be homogeneous of degree s with coefficients in A. If*

$$f(x_1,\dots,x_d)\in\mathfrak{q}^{s+1},$$

then every coefficient of f belongs to $\mathfrak{m}$.

Proof. For $s=0$ the assertion simply says that a constant in $\mathfrak{q}$ belongs to $\mathfrak{m}$. Thus suppose $s>0$ and, seeking a contradiction, assume that some coefficient is outside $\mathfrak{m}$, hence is a unit. Put $B=A/\mathfrak{q}$, an Artinian local ring, and reduce f to a polynomial $\bar f\in R=B[t_1,\dots,t_d]$. Its unit coefficient stays a unit, so Lemma 12.21 says it is a non-zero-divisor in R.

There is a surjective graded map

$$R\longrightarrow G_{\mathfrak{q}}(A),\qquad t_i\longmapsto x_i \bmod \mathfrak{q}^2.$$

Its degree-s value on $\bar f$ is the class of $f(x_1,\dots,x_d)$ modulo $\mathfrak{q}^{s+1}$, which is zero. The map therefore factors through a surjection $R/(\bar f)\to G_{\mathfrak{q}}(A)$.

Compute Hilbert series using length over B. Multiplication by the homogeneous non-zero-divisor $\bar f$ yields

$$H_{R/(\bar f)}(t)=\frac{\operatorname{length}(B)(1-t^s)}{(1-t)^d}.$$

It has pole order $d-1$, since $1-t^s$ has a simple zero at one. The cumulative sum of its coefficients is consequently eventually a polynomial of degree $d-1$; for $d=1$ this is an eventually positive constant. Surjectivity in every degree bounds the cumulative lengths for $G_{\mathfrak{q}}(A)$ by these cumulative lengths. But the cumulative lengths for $G_{\mathfrak{q}}(A)$ are $\operatorname{length}(A/\mathfrak{q}^n)$, whose polynomial degree is $\dim A=d$ by Theorem 12.14. A positive polynomial of degree d cannot be bounded by one of degree $d-1$. This contradiction proves that every coefficient was in $\mathfrak{m}$. If $d=0$, a polynomial in the empty list of variables is constant, already covered by the first case. □

A *coefficient field* of $(A,\mathfrak{m})$ is a subfield $k\subseteq A$ for which the residue map $k\to A/\mathfrak{m}$ is an isomorphism. It allows each residue class to be represented by a unique element of this subfield. For example, the constants form a coefficient field in $k[t]_{(t)}$. Such a field need not exist: $\mathbb{Z}_{(p)}$ has characteristic zero whereas its residue field has characteristic p.

Corollary 12.23 (Parameters over a coefficient field; AM 11.21)**.** *Suppose a subfield $k\subseteq A$ maps isomorphically onto $A/\mathfrak{m}$. Then any system of parameters $x_1,\dots,x_d$ is algebraically independent over k.*

Proof. Suppose $0\neq f\in k[t_1,\dots,t_d]$ satisfies $f(x_1,\dots,x_d)=0$. Decompose f into homogeneous pieces, and let f_s be its nonzero piece of smallest degree. Every higher-degree piece evaluates into $\mathfrak{q}^{s+1}$, so the relation implies $f_s(x_1,\dots,x_d)\in\mathfrak{q}^{s+1}$. Proposition 12.22 puts every coefficient of f_s in $\mathfrak{m}$. But $k\cap\mathfrak{m}=0$, because nonzero elements of the field are units in A. Hence all those coefficients are zero, contradicting the choice of f_s. □

Parameters need not form a regular sequence

A sequence $a_1,\dots,a_r$ in a local ring is called a regular sequence if its ideal is proper and multiplication by a_i is injective on $A/(a_1,\dots,a_{i-1})$ for each i. This condition is stronger than being a system of parameters.

For example, take

$$A = \big(k[x,y]/(x^2,xy)\big)_{(x,y)}, \qquad \mathfrak{m} = (x,y)A.$$

The nilradical is (x) and $A/(x) \cong k[y]_{(y)}$, so $\dim A = 1$. The quotient

$$A/(y) \cong k[x]/(x^2)$$

has finite length. Hence y is a system of parameters. But $yx = 0$ and $x \neq 0$, so y is a zero divisor and is not a regular sequence of length one. The nonzero class of x is seen already from the monomial basis before localization, and cannot be killed by a polynomial with nonzero constant term.

This example also separates dimension from the number of generators of the maximal ideal. The two classes x, y are independent in $\mathfrak{m}/\mathfrak{m}^2$, so

$$\dim A = 1 < 2 = \dim_k \mathfrak{m}/\mathfrak{m}^2.$$

For $n \geq 2$, a basis of $A/\mathfrak{m}^n$ is $1, y, \dots, y^{n-1}, x$. Its length is $n+1$, of degree one in n, as the dimension theorem predicts. The extra nilpotent contributes one to the constant term without increasing dimension.

12.7 Regular local rings and tangent spaces

Theorem 12.24 (Equivalent tests for regularity; AM 11.22)**.** *Let $(A,\mathfrak{m},\kappa)$ be Noetherian local of dimension d. The following conditions are equivalent:*

1. *$G_{\mathfrak{m}}(A)$ is isomorphic, as a graded κ-algebra, to a polynomial ring $\kappa[t_1,\dots,t_d]$ with all variables of degree one;*
2. *$\dim_\kappa(\mathfrak{m}/\mathfrak{m}^2) = d$;*
3. *$\mathfrak{m}$ can be generated by d elements.*

A ring satisfying them is called a regular local ring.

Proof. The degree-one component in (i) has basis $t_1,\dots,t_d$, proving (ii). Lifting a basis in (ii) and applying Nakayama gives (iii). Finally suppose $\mathfrak{m} = (x_1,\dots,x_d)$. This list is a system of parameters, and the degree-one classes induce a surjective graded map $\kappa[t_1,\dots,t_d] \to G_{\mathfrak{m}}(A)$. Its kernel is homogeneous. If a homogeneous polynomial of degree s belongs to the kernel, lift its coefficients to A. The resulting polynomial f satisfies $f(x_1,\dots,x_d) \in \mathfrak{m}^{s+1}$. Proposition 12.22 says all its coefficients are in $\mathfrak{m}$, so the original

polynomial over κ is zero. The kernel has no nonzero homogeneous elements and is therefore zero. This proves (i). □

Lemma 12.25 (A domain criterion from the associated graded ring; AM 11.23)**.** *Let A be a ring and I an ideal with $\bigcap_{n\geq 0} I^n = 0$. If $G_I(A)$ is a domain, then A is a domain.*

Proof. Since $G_I(A) \neq 0$, also $A \neq 0$. For nonzero $x, y \in A$, the zero intersection gives finite integers $r, s \geq 0$ with

$$x \in I^r \setminus I^{r+1}, \qquad y \in I^s \setminus I^{s+1}.$$

Their classes $\bar{x} \in I^r/I^{r+1}$ and $\bar{y} \in I^s/I^{s+1}$ are nonzero. The domain hypothesis implies

$$0 \neq \bar{x}\bar{y} = xy + I^{r+s+1} \in I^{r+s}/I^{r+s+1}.$$

Thus $xy \notin I^{r+s+1}$, and in particular $xy \neq 0$. □

Every regular local ring is consequently a domain: its associated graded ring is a polynomial domain, and Corollary 11.19 gives the required zero intersection. In dimension one the regular local rings are precisely the DVRs of Lecture 10. The ring $\mathbb{Z}_{(p)}$ is a useful example; regularity does not require that the ring contain a field.

The vector space $\mathfrak{m}/\mathfrak{m}^2$ is called the *cotangent space* of the local ring. Its dimension $\operatorname{edim} A = \dim_\kappa(\mathfrak{m}/\mathfrak{m}^2)$ is the *embedding dimension*; it is at least the Krull dimension by Corollary 12.16. The *dual* of a vector space W over κ is the vector space $W^* = \operatorname{Hom}_\kappa(W, \kappa)$ of all linear functionals $W \to \kappa$. Accordingly, the *tangent space* of $(A, \mathfrak{m})$ is

$$T_A = \operatorname{Hom}_\kappa(\mathfrak{m}/\mathfrak{m}^2, \kappa).$$

For finite-dimensional W, choosing a basis and its coordinate functionals shows that $\dim W^* = \dim W$.

To interpret tangent vectors, recall the ring of *dual numbers* $k[\varepsilon]/(\varepsilon^2)$. Each element is uniquely $a + b\varepsilon$, with $a, b \in k$, and multiplication uses $\varepsilon^2 = 0$. Its residue map is $a + b\varepsilon \mapsto a$, and an element is a unit exactly when $a \neq 0$, with inverse $a^{-1} - a^{-2}b\varepsilon$.

Suppose now that A has a coefficient field k, and write the residue of a as $a(p) \in k$. The symbol p names the point under consideration; only the residue map is needed in the following calculation. A k-algebra map $A \to k[\varepsilon]/(\varepsilon^2)$ reducing to the residue map must have the form $\phi(a) = a(p) + D(a)\varepsilon$. Additivity and the condition on k say that $D : A \to k$ is k-linear and vanishes on k. Multiplicativity is precisely the *Leibniz rule*

$$D(ab) = a(p)D(b) + b(p)D(a).$$

A map with these properties is called a *k-derivation at p*. If $a, b \in \mathfrak{m}$, their residues vanish, so the rule gives $D(ab) = 0$. Therefore D kills $\mathfrak{m}^2$ and induces a linear functional on $\mathfrak{m}/\mathfrak{m}^2$. Conversely, a functional λ defines $D(a) = \lambda(a - a(p) \bmod \mathfrak{m}^2)$. Expanding a product verifies the displayed identity, because the product of two elements of $\mathfrak{m}$ disappears modulo $\mathfrak{m}^2$. This identifies tangent vectors with maps into the dual numbers, retaining exactly first-order information.

For a quotient $R = k[x_1,\dots,x_r]/(f_1,\dots,f_s)$, a *k-rational point* is a tuple $p = (p_1,\dots,p_r) \in k^r$ satisfying every equation $f_i(p) = 0$. Evaluation is then a surjective map $R \to k$. Its kernel is the maximal ideal $\mathfrak{m}_p = (x_1 - p_1,\dots,x_r - p_r)$, where the generators denote their classes in R. The *local ring at p* is $R_{\mathfrak{m}_p}$, and the tangent space at p means $T_{R_{\mathfrak{m}_p}}$ as defined above.

We use *formal partial derivatives*: for a monomial,

$$\frac{\partial}{\partial x_j}(x_1^{a_1}\cdots x_r^{a_r}) = a_j x_1^{a_1}\cdots x_j^{a_j-1}\cdots x_r^{a_r},$$

and extend by k-linearity; when $a_j = 0$ the derivative is zero. The integer a_j is read as an element of k, so the definition also applies in positive characteristic.

Proposition 12.26 (Tangent vectors from equations)**.** *Let the ring of functions be*

$$R = k[x_1,\dots,x_r]/(f_1,\dots,f_s),$$

and let p be a k-rational point. Its tangent vectors are tuples $v \in k^r$ satisfying

$$\sum_{j=1}^{r} \frac{\partial f_i}{\partial x_j}(p)\, v_j = 0 \qquad (1 \le i \le s).$$

The same tangent space is obtained after localizing at the point.

Proof. A homomorphism to the dual numbers reducing to evaluation at p has

$$x_j \longmapsto p_j + v_j\varepsilon, \qquad \varepsilon^2 = 0.$$

Expanding each monomial gives

$$f_i(p + v\varepsilon) = f_i(p) + \varepsilon \sum_{j=1}^{r} \frac{\partial f_i}{\partial x_j}(p)\, v_j.$$

The constant term is zero because p lies on the zero set. Hence

$$f_i(p + v\varepsilon) = 0 \text{ for every } i \quad\Longleftrightarrow\quad \sum_j \frac{\partial f_i}{\partial x_j}(p)\, v_j = 0 \text{ for every } i.$$

Killing the generators f_i kills their ideal, so these are exactly the conditions for the map to factor through R. Every denominator allowed in $R_{\mathfrak{m}_p}$ has nonzero residue; its image in the dual numbers is therefore a unit. The map extends uniquely to the local ring. □

We call a k-rational point *nonsingular*, or *regular*, when its local ring is regular, and *singular* otherwise. Over an algebraically closed field this is the usual nonsingularity condition for an affine variety; the equation test below is proved directly from the local ring.

Example 12.27 (A nonsingular graph and a cusp)**.** For $y - x^2 = 0$ at the origin, the tangent equation is $v_y = 0$. Substituting $y = x^2$ identifies the local ring with $k[x]_{(x)}$, which is regular of dimension one.

For $y^2 - x^3 = 0$, both partial derivatives vanish at the origin in every characteristic. The tangent space is two-dimensional, while the local dimension is one by Corollary 12.19, applied to the nonzero equation in $k[x,y]_{(x,y)}$. Thus the cusp is singular: it has more first-order directions than its dimension.

Deriving the tangent-space equations from dual numbers

Let $X \subseteq \mathbb{A}^n_k$ be defined by an ideal I, and let $p = (a_1, \dots, a_n) \in X(k)$. A first-order displacement is a tuple

$$a_i + \varepsilon v_i \in k[\varepsilon]/(\varepsilon^2).$$

Substitute it into a polynomial f. Expanding monomials and discarding all terms containing ε^2 gives

$$f(a + \varepsilon v) = f(a) + \varepsilon \sum_i \frac{\partial f}{\partial x_i}(a) v_i.$$

Since $f(a) = 0$ for $f \in I$, the displacement remains on X to first order exactly when

$$\sum_i f_{x_i}(p) v_i = 0 \qquad (f \in I).$$

It suffices to use generators of I: the product rule shows that the differential of $\sum_j h_j f_j$ at p equals $\sum_j h_j(p)\, df_j(p)$.

To compare this description with $\mathrm{Hom}_k(\mathfrak{m}_p/\mathfrak{m}_p^2, k)$, assign to v the map sending the class of $x_i - a_i$ to v_i. Products of two functions vanishing at p must be sent to zero because their evaluation at $a + \varepsilon v$ has order at least two. Relations in I impose precisely the displayed linear equations. Thus this assignment gives the intrinsic identification

$$T_p X = \mathrm{Hom}_k(\mathfrak{m}_p/\mathfrak{m}_p^2, k).$$

The vector space $\mathfrak{m}_p/\mathfrak{m}_p^2$ is the cotangent space; its dual is the tangent space. Their dimensions agree, but the spaces have different variance under maps.

12.8 Transcendence degree and the dimension of a variety

We first recall the distinction between algebraic dependence and linear dependence. Elements $u_1, \dots, u_d$ in an extension field L of k are *algebraically independent over k* if no nonzero polynomial in $k[T_1, \dots, T_d]$ vanishes on the tuple. Equivalently, the evaluation homomorphism $k[T_1, \dots, T_d] \to L$ is injective. Its image is then a polynomial ring $k[u_1, \dots, u_d]$, whose fraction field is denoted by $k(u_1, \dots, u_d)$.

Definition 12.28. A *transcendence basis* for L/k is an algebraically independent tuple $u_1,\dots,u_d$ such that every element of L is algebraic over $k(u_1,\dots,u_d)$. We shall only need extensions that are finitely generated as fields, so the tuple can be taken finite. A field extension generated by the empty tuple means the original field k.

We will use transitivity of algebraicity. If L/K and K/F are algebraic and $\alpha \in L$, choose a monic equation for α over K. Its finitely many coefficients generate a finite extension F'/F, by the tower formula. Then $F'(\alpha)/F'$ is finite, so the tower formula makes $F'(\alpha)/F$ finite as well. Hence α is algebraic over F.

To construct a transcendence basis, take a finite list of field generators and choose a maximal algebraically independent sublist. Every omitted generator is algebraic over the field generated by the chosen ones. Indeed, a polynomial dependence with that additional generator gives a nonzero polynomial in it over this field, since the chosen generators themselves have no relation. Adjoining the omitted algebraic generators makes the whole extension algebraic. We next prove that the length of the tuple is independent of the choice.

Lemma 12.29 (Exchange lemma)**.** *Suppose L is algebraic over $F(u_1,\dots,u_r)$ and $v \in L$ is transcendental over F. Then one of the u_j can be replaced by v so that L is still algebraic over the field generated by the resulting list and F.*

Proof. The element v is algebraic over $F(u_1,\dots,u_r)$, but not over F. Choose the least j for which it is algebraic over $F(u_1,\dots,u_j)$, and put $E = F(u_1,\dots,u_{j-1})$. Clear the denominators in a nonzero polynomial relation for v over $E(u_j)$. This gives

$$P(u_j,v) = 0, \qquad 0 \neq P(X,Y) \in E[X,Y],$$

with the polynomial $P(u_j,Y)$ nonzero. Since v is transcendental over E, substitution $Y = v$ is injective on $E[Y]$, so $P(X,v)$ is a nonzero polynomial over $E(v)$. It cannot be a nonzero constant, because its value at u_j is zero. Thus u_j is algebraic over $E(v)$.

Let F' be the field generated by F, v, and all the u_i except u_j. Then u_j is algebraic over F'. The field L is algebraic over $F'(u_j)$ because it was algebraic over the smaller field $F(u_1,\dots,u_r)$. Transitivity of algebraic extensions therefore shows that L is algebraic over F', as required. □

Proposition 12.30 (Invariance of the size of a transcendence basis)**.** *If L is algebraic over $k(u_1,\dots,u_r)$, every algebraically independent tuple in L over k has at most r elements. Consequently any two finite transcendence bases have the same number of elements.*

Proof. Let $v_1,\dots,v_s$ be algebraically independent and suppose, for a contradiction, that $s > r$. If $r = 0$, the extension L/k is algebraic, which already contradicts the transcendence of v_1. Thus assume $r > 0$. The exchange lemma inserts v_1 into the given list, deleting one u_i and preserving algebraicity of L over the generated field.

If $r > 1$, the element v_2 is transcendental over $k(v_1)$: an algebraic relation would, after clearing denominators, contradict independence of v_1, v_2. Apply the lemma again, now with base field $k(v_1)$, to insert v_2 and delete another remaining u_i. Repeating this argument inserts the first r elements of the tuple. It follows that L is algebraic over $k(v_1,\dots,v_r)$. Since $s > r$, the element v_{r+1} would then be algebraic over that field, contradicting independence.

Hence $s \le r$. Applying this inequality in both directions to two transcendence bases proves equality of their sizes. □

Definition 12.31. The common size of a transcendence basis is the *transcendence degree*, written $\operatorname{trdeg}_k L$.

For example, the indeterminates x, y form a transcendence basis of $k(x,y)$, so its transcendence degree is two. The field $k(t^2,t^3)$ equals $k(t)$, because $t = t^3/t^2$; its transcendence degree is one. If L is algebraic over an intermediate field K containing k, a transcendence basis for K/k is also one for L/k, by transitivity of algebraicity. Thus algebraic extension does not change the transcendence degree.

For clarity, the geometric terminology needed for the next theorem is introduced here. Let k be algebraically closed. The *affine r-space* is $\mathbb{A}^r_k = k^r$. An *affine algebraic set* is the set V of common zeros in k^r of a family of polynomials. We write $V(f)$ for the zeros of a single polynomial f, and $V(f_1,\dots,f_s)$ for the common zeros of a finite list. A set $V(f)$ for a nonzero nonconstant polynomial f is an *affine hypersurface*. For any affine algebraic set V, define its *vanishing ideal* and *coordinate ring* by

$$I(V) = \{f \in k[x_1,\dots,x_r] : f(p) = 0 \text{ for every } p \in V\}, \qquad k[V] = k[x_1,\dots,x_r]/I(V).$$

Two polynomials give the same function on V exactly when their difference lies in $I(V)$. Thus the coordinate ring is precisely the ring of polynomial functions on V.

A subset of V obtained by imposing further polynomial equations is called *closed in* V. A nonempty algebraic set V is *irreducible* if it is not the union of two proper closed subsets. An *irreducible affine variety* means a nonempty irreducible affine algebraic set. Its coordinate ring is a domain. In fact, if fg vanishes on V, then $V = (V \cap V(f)) \cup (V \cap V(g))$, so irreducibility forces one of f, g to vanish on all of V. This is the prime-ideal condition for $I(V)$. Conversely, if $V = V_1 \cup V_2$ with both closed subsets proper, choose a polynomial vanishing on V_1 but not on all of V, and one vanishing on V_2 but not on all of V. Such polynomials exist by the definition of the closed subsets and their properness. Their product vanishes on V, so $I(V)$ is not prime. This proves the equivalence.

For an irreducible affine variety V, the *field of rational functions* is $k(V) = \operatorname{Frac}(k[V])$, and its classical dimension is defined as $\operatorname{trdeg}_k k(V)$. A rational function f/g can be evaluated at a point where g does not vanish. Evaluation at $p \in V$ gives the maximal ideal $\mathfrak{m}_p$ defined in the tangent-space discussion, and $k[V]_{\mathfrak{m}_p}$ consists of the fractions having a denominator nonzero at p. The weak Nullstellensatz identifies all maximal ideals with such points: applied to the polynomial ring it writes their preimages as $(x_1 - p_1,\dots,x_r - p_r)$; containment of $I(V)$ puts p in the original algebraic set. The next theorem compares this geometric dimension with the local Krull dimension.

Lemma 12.32 (Local dimension under an integral extension; AM 11.26)**.** *Let $B \subseteq R$ be domains, with B integrally closed and R integral over B. If $\mathfrak{m}$ is maximal in R and $\mathfrak{n} = \mathfrak{m} \cap B$, then $\mathfrak{n}$ is maximal and $\dim R_{\mathfrak{m}} = \dim B_{\mathfrak{n}}$. No Noetherian assumption is needed for this statement.*

Proof. The field criterion for integral extensions (AM 5.8) shows that contraction of a maximal ideal is maximal. Now take a strict chain of primes in R ending at $\mathfrak{m}$. Contracting it

to B produces a strict chain ending at $\mathfrak{n}$: if two comparable primes had the same contraction, incomparability (AM 5.9) would force them to be equal. Every length counted for $R_{\mathfrak{m}}$ is therefore counted for $B_{\mathfrak{n}}$, proving the inequality $\leq$.

Conversely, take any finite strict prime chain in B ending at $\mathfrak{n}$. Start with the chosen prime $\mathfrak{m}$ above its final term. The going-down theorem (AM 5.16), whose normal-domain hypotheses hold here, lifts the preceding prime to one contained in $\mathfrak{m}$. Repeat down the chain. Distinct contractions ensure distinct lifted primes, giving a strict chain of the same length ending at $\mathfrak{m}$. This proves the reverse inequality after taking suprema. Going-down is the reason this argument can work below a prescribed maximal ideal upstairs. □

Theorem 12.33 (Local dimension equals transcendence degree; AM 11.25)**.** *For an irreducible affine variety V over an algebraically closed field k, every point, with corresponding maximal ideal $\mathfrak{m} \subseteq R = k[V]$, satisfies*

$$\dim R_{\mathfrak{m}} = \operatorname{trdeg}_k \operatorname{Frac}(R) = \dim V.$$

Proof. Let $d = \operatorname{trdeg}_k \operatorname{Frac}(R)$. Noether normalization, Theorem 10.37, supplies algebraically independent elements $u_1, \dots, u_d \in R$ such that $B = k[u_1, \dots, u_d] \subseteq R$ is finite and integral. The number is d because integrality makes $\operatorname{Frac}(R)$ algebraic over $k(u_1, \dots, u_d)$.

A polynomial ring over a field is a unique factorization domain by Theorem 10.11, and hence is integrally closed by Proposition 10.12. Thus Lemma 12.32 gives $\dim R_{\mathfrak{m}} = \dim B_{\mathfrak{n}}$ for $\mathfrak{n} = \mathfrak{m} \cap B$. Since k is algebraically closed, the weak Nullstellensatz (Corollary 9.24) says $\mathfrak{n} = (u_1 - a_1, \dots, u_d - a_d)$ with $a_i \in k$. Translate coordinates by setting $v_i = u_i - a_i$. The resulting local ring is $k[v_1, \dots, v_d]_{(v_1, \dots, v_d)}$. Its associated graded monomial calculation above, together with Theorem 12.14, gives dimension d. This proves the claimed equality at the original point. □

Corollary 12.34 (Global dimension agrees with every closed local dimension; AM 11.27)**.** *For every maximal ideal $\mathfrak{m}$ of the coordinate domain $R = k[V]$ above,*

$$\dim R = \dim R_{\mathfrak{m}} = \operatorname{trdeg}_k \operatorname{Frac}(R).$$

Proof. Proposition 12.2 identifies $\dim R$ with the supremum of $\dim R_{\mathfrak{m}}$ over maximal ideals. The preceding theorem makes all these numbers equal to the same finite integer d, so their supremum is d. □

12.9 Worked Hilbert–Samuel calculations

Some authors, and our examples below, index the Hilbert–Samuel function as $L_A(n) = \operatorname{length}(A/\mathfrak{m}^{n+1})$ for $n \geq 0$. This is the shift $\chi_{\mathfrak{m}}(n+1)$ for large n. Shifting the argument leaves both degree and leading coefficient unchanged.

For a line, $L_A(n) = n + 1$. For a plane at the origin, counting monomials of total degree at most n gives $L_A(n) = \binom{n+2}{2}$. For the union of the two coordinate axes, whose crossing at the origin is called a *node*, let $A = (k[x,y]/(xy))_{(x,y)}$. A basis modulo $\mathfrak{m}^{n+1}$ is

$$1, x, \dots, x^n, y, \dots, y^n,$$

so $L_A(n) = 2n+1$. For the ring $B = (k[x,y]/(x^2))_{(x,y)}$, called a *double line* because its defining equation is the square of the equation $x = 0$, a basis is $1, y, \dots, y^n$ together with $x, xy, \dots, xy^{n-1}$, again giving $2n+1$. The first ring is reduced and records the two coordinate axes; the second has the nonzero nilpotent x. These growth functions consequently do not classify local rings.

If $d = \dim A$, write the leading coefficient of $L_A(n)$ as $e(A)/d!$. The number $e(A)$ is the *multiplicity*. It is a positive integer. To see integrality, define the *forward difference* of a polynomial by $\Delta P(n) = P(n+1) - P(n)$. The binomial formula shows that Δ lowers a positive degree by one and multiplies its leading coefficient by that degree. Thus $\Delta^d P$ is the constant $d!$ times the leading coefficient. If $P(n)$ is an integer for every sufficiently large integer n, then the same is true of every iterated difference, so this constant is an integer. Positivity follows because the length polynomial is eventually positive. Thus a line has multiplicity one, while the node and double line above have multiplicity two.

12.10 Exercises

Exercise 12.1 (A parameter on the cusp)**.** For $A = (k[x,y]/(y^2 - x^3))_{(x,y)}$ with maximal ideal $\mathfrak{m} = (x,y)A$, show that x is a system of parameters, although $\mathfrak{m}$ needs two generators. Compute $\operatorname{length} A/(x)$.

Exercise 12.2 (Three rings with different dimensions)**.** Compute the dimensions of $k[t^2, t^3]$, $k[x,y]/(xy)$, and $k[x,y]/(x^2, y^3)$, first taking k algebraically closed.

Exercise 12.3 (A thickened point)**.** Let $A = k[t]_{(t)}/(t^r)$ with $r \geq 1$ and maximal ideal $\mathfrak{m} = (t)$. Compute $L_A(n) = \operatorname{length}_A(A/\mathfrak{m}^{n+1})$ for $n \geq 0$, and determine $\dim A$.

Exercise 12.4 (The hypersurface singularity test; AM Chapter 11, Exercise 1)**.** Let k be algebraically closed and let $f \in k[x_1, \dots, x_n]$ be irreducible. For a point P on $f = 0$, prove that the local ring of $k[x_1, \dots, x_n]/(f)$ at P is regular if and only if at least one partial derivative $\partial f/\partial x_i$ is nonzero at P.

Exercise 12.5 (The ground field need not be algebraically closed; AM Chapter 11, Exercise 3)**.** Let R be a domain of finite type over an arbitrary field k. Prove $\dim R = \operatorname{trdeg}_k \operatorname{Frac}(R)$. Show also that every maximal ideal $\mathfrak{m} \subseteq R$ satisfies $\dim R_{\mathfrak{m}} = \operatorname{trdeg}_k \operatorname{Frac}(R)$.

Exercise 12.6 (Polynomial dimension without Noetherianity; AM Chapter 11, Exercise 6)**.** For a nonzero ring A, prove

$$1 + \dim A \leq \dim A[t] \leq 1 + 2\dim A,$$

with infinite dimensions interpreted as ∞.

Exercise 12.7 (Polynomial dimension in the Noetherian case; AM Chapter 11, Exercise 7)**.** If A is nonzero and Noetherian, prove

$$\dim A[t] = \dim A + 1, \qquad \dim A[t_1, \dots, t_r] = \dim A + r.$$

Exercise 12.8 (Hilbert functions and dimension)**.** Give $A = k[x, y]/(x^2)$ its standard grading. Compute its Hilbert function and Hilbert series, its Krull dimension, and the lengths of $R/\mathfrak{m}^{n+1}$ for $R = A_{(x,y)}$. Is this local ring regular?

Lecture 13
Categories, functors, and complexes

Categories organize objects together with their structure-preserving maps. Complexes and cohomology provide a way to measure the failure of a sequence to be exact. This lecture introduces the language needed for cohomology. Reading: [2], Chapter III, Section 1, for the homological algebra vocabulary; our discussion supplies the elementary explanations that the reference presupposes.

13.1 Categories and functors

A vector space is studied together with linear maps, a group together with group homomorphisms, and a ring together with ring homomorphisms. Category theory records this recurring organization. It does not replace calculations; it explains which calculations can be transported from one setting to another.

Definition 13.1. A *category* $\mathscr{C}$ consists of objects, sets $\mathrm{Hom}_{\mathscr{C}}(X,Y)$ of morphisms for each ordered pair of objects, identity morphisms id_X, and composition maps

$$\mathrm{Hom}_{\mathscr{C}}(Y,Z)\times\mathrm{Hom}_{\mathscr{C}}(X,Y)\longrightarrow\mathrm{Hom}_{\mathscr{C}}(X,Z),\qquad (g,f)\longmapsto g\circ f.$$

For composable morphisms, the axioms are

$$h\circ(g\circ f)=(h\circ g)\circ f,\qquad \mathrm{id}_Y\circ f=f=f\circ\mathrm{id}_X\,.$$

A morphism $f:X\to Y$ is an *isomorphism* if there is a morphism $g:Y\to X$ with

$$g\circ f=\mathrm{id}_X,\qquad f\circ g=\mathrm{id}_Y\,.$$

The category R-Mod has R-modules as objects and R-linear maps as morphisms. The category Set has sets and functions. The category k-Alg has commutative k-algebras and homomorphisms commuting with the specified structure map from k. A ring map need not be a k-algebra map: complex conjugation preserves the ring structure on $\mathbb{C}$ but does not fix every complex scalar.

An isomorphism depends on the category. The same underlying bijection may preserve one structure and fail to preserve another. Conversely, the inverse of a bijective linear map is linear: if $T(v)=w$ and $T(v')=w'$, then $T(av+bv')=aw+bw'$, which proves the assertion by applying T^{-1}. Later we will meet bijective geometric morphisms whose inverses are not morphisms.

A diagram consists of objects and specified arrows. It is *commutative* if paths with the same endpoints have equal composites. For example,

$$\begin{array}{ccc} X & \xrightarrow{\ f\ } & Y \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ X' & \xrightarrow[\ g\ ]{} & Y' \end{array} \qquad \Longleftrightarrow \qquad v\circ f = g\circ u.$$

For modules, this means $v(f(x)) = g(u(x))$ for every $x \in X$.

Definition 13.2. A covariant *functor* $F : \mathscr{C} \to \mathscr{D}$ assigns an object $F(X)$ to each object and a morphism $F(f) : F(X) \to F(Y)$ to each morphism $f : X \to Y$, with

$$F(\mathrm{id}_X) = \mathrm{id}_{F(X)}, \qquad F(g\circ f) = F(g)\circ F(f).$$

A contravariant functor reverses arrows and satisfies $F(g\circ f) = F(f)\circ F(g)$.

Forgetting scalar multiplication gives a functor from R-modules to abelian groups. For a fixed module P and a map $u : M \to N$, the two Hom functors act as follows:

$$\begin{aligned} \mathrm{Hom}_R(P,M) &\longrightarrow \mathrm{Hom}_R(P,N), \quad h \longmapsto u\circ h, \\ \mathrm{Hom}_R(N,P) &\longrightarrow \mathrm{Hom}_R(M,P), \quad h \longmapsto h\circ u. \end{aligned}$$

The first is covariant; the second is contravariant.

Tensoring with a fixed module P is covariant, sending u to the map $\mathrm{id}_P \otimes u$. The formula is defined on elementary tensors by $p\otimes m \mapsto p\otimes u(m)$; the balancing relation makes it well-defined. An equality of composites can then be checked on elementary tensors, since they generate the tensor product.

13.2 Natural transformations and universal properties

Definition 13.3. For two covariant functors $F, G : \mathscr{C} \to \mathscr{D}$, a *natural transformation* $\eta : F \Rightarrow G$ is a family of morphisms $\eta_X : F(X) \to G(X)$ such that

$$G(f)\circ \eta_X = \eta_Y \circ F(f) \quad \text{for every } f : X \to Y.$$

It is a natural isomorphism if every η_X is an isomorphism.

"Natural" means compatible with all the allowed maps. It does not merely mean that a formula looks convenient. For example,

$$\eta_M : R \otimes_R M \longrightarrow M, \qquad r \otimes m \longmapsto rm$$

is a natural isomorphism. Its inverse sends m to $1 \otimes m$. For a linear map $u : M \to N$, both routes around the naturality square send $r \otimes m$ to $ru(m)$.

The dual vector space is $V^* = \mathrm{Hom}_k(V,k)$. A linear map $T : V \to W$ induces $T^* : W^* \to V^*$ by $T^*(\mu) = \mu \circ T$. Dualizing twice gives $T^{**} : V^{**} \to W^{**}$, so the double dual construction is covariant. There is a natural map

$$\varepsilon_V : V \longrightarrow V^{**}, \qquad \varepsilon_V(v)(\lambda) = \lambda(v).$$

For $T : V \to W$, $v \in V$, and $\mu \in W^*$,

$$\begin{aligned} \big(T^{**}\varepsilon_V(v)\big)(\mu) &= \varepsilon_V(v)(\mu \circ T) \\ &= \mu(Tv) = \varepsilon_W(Tv)(\mu). \end{aligned}$$

Thus $T^{**}\varepsilon_V = \varepsilon_W T$. In finite dimension ε_V is an isomorphism. A basis also gives $V \cong V^*$, but that construction depends on the chosen basis.

Definition 13.4. Functors $F : \mathscr{C} \to \mathscr{D}$ and $G : \mathscr{D} \to \mathscr{C}$ are *adjoint*, with F left adjoint to G, if there are bijections

$$\mathrm{Hom}_{\mathscr{D}}(F(X), Y) \cong \mathrm{Hom}_{\mathscr{C}}(X, G(Y))$$

compatible with composition in both X and Y. This compatibility is called naturality in the two variables. It means that applying a map before or after either side gives the corresponding map on the other side.

For example, the tensor–Hom bijection from Lecture 5 makes $P \otimes_R -$ left adjoint to $\mathrm{Hom}_R(P,-)$. It sends $h : P \otimes_R M \to N$ to the map $m \mapsto (p \mapsto h(p \otimes m))$. Its inverse evaluates this formula on elementary tensors. Thus an adjunction records a previously proved correspondence between concrete linear maps.

For $f : M \to N$, write $i : \ker f \hookrightarrow M$ and $q : N \to N/\operatorname{im} f$. Their universal properties are

$$\begin{aligned} fg = 0 &\implies \exists!\, \widetilde{g} : P \to \ker f, \qquad i\widetilde{g} = g, \\ hf = 0 &\implies \exists!\, \widetilde{h} : N/\operatorname{im} f \to Q, \quad \widetilde{h}q = h. \end{aligned}$$

Explicitly, $\widetilde{g}(p) = g(p)$ and $\widetilde{h}(n + \operatorname{im} f) = h(n)$. The first formula lands in the kernel because $fg = 0$; the second is independent of representatives because $hf = 0$.

An object 0 is *initial* if there is exactly one morphism $0 \to X$ for each X, and *terminal* if there is exactly one morphism $X \to 0$ for each X. A *zero object* has both properties; the zero module is an example. The composite $X \to 0 \to Y$ is then the *zero morphism* from X to Y. With this meaning of zero, the universal descriptions of kernels and cokernels above apply in any category with a zero object.

A *product* of objects X, Y is an object P with projections $p_X : P \to X$, $p_Y : P \to Y$ such that maps $a : T \to X$, $b : T \to Y$ uniquely determine a map $T \to P$ whose composites with the projections are a, b. A *coproduct* reverses all these arrows: maps from X, Y to a common object uniquely extend from the coproduct. In modules, $X \oplus Y$ serves both purposes. When morphisms can be added, a *biproduct* has inclusions i_X, i_Y and projections satisfying

$$p_X i_X = \mathrm{id}_X, \quad p_Y i_Y = \mathrm{id}_Y, \quad p_X i_Y = p_Y i_X = 0, \qquad i_X p_X + i_Y p_Y = \mathrm{id}.$$

These identities express the two universal properties without referring to elements.

Definition 13.5. An *abelian category* is a category in which morphism sets are abelian groups, composition is additive in each variable, a zero object and finite biproducts exist, every morphism has a kernel and cokernel, and the canonical map

$$\operatorname{coim} f := \operatorname{coker}(\ker f \to X) \longrightarrow \operatorname{im} f := \ker(Y \to \operatorname{coker} f)$$

is an isomorphism for every $f : X \to Y$.

The map from coimage to image in the definition is the map induced by f after its kernel has been removed from the source and its image has been selected in the target. Its construction uses the two universal factorizations above. In modules its formula is $x + \ker f \mapsto f(x)$.

Finite direct sums here are also categorical products; such objects are called *biproducts*. In modules, $M \oplus N$ has coordinate projections and coordinate inclusions, and the identity is the sum of the two inclusion–projection composites. The last axiom of an abelian category is precisely the familiar first isomorphism theorem

$$M/\ker f \cong \operatorname{im} f.$$

Thus R-Mod and the category of abelian groups are abelian. Vector spaces give another example. The category whose objects are free $\mathbb{Z}$-modules and whose morphisms are all homomorphisms between them is not abelian. For multiplication by 2 on $\mathbb{Z}$, the kernel is zero. Its cokernel *within this category* is also zero: if $h : \mathbb{Z} \to F$ with F free satisfies $h(2) = 0$, then $2h(1) = 0$, whence $h(1) = 0$. The canonical map from coimage to image is consequently multiplication by 2 from $\mathbb{Z}$ to $\mathbb{Z}$, which is not an isomorphism. In the larger category of all abelian groups, the cokernel is instead $\mathbb{Z}/2\mathbb{Z}$. The category in which a universal property is tested matters.

13.3 Complexes and cohomology

Definition 13.6. A *cochain complex* of R-modules is a sequence

$$\cdots \longrightarrow C^{i-1} \xrightarrow{d^{i-1}} C^i \xrightarrow{d^i} C^{i+1} \longrightarrow \cdots$$

with $d^i d^{i-1} = 0$ for every i. Its cycles, boundaries, and cohomology are

$$Z^i(C) = \ker d^i, \qquad B^i(C) = \operatorname{im} d^{i-1}, \qquad H^i(C) = Z^i(C)/B^i(C).$$

Unspecified terms of a finite displayed complex are understood to be zero.

The condition $d^2 = 0$ gives $B^i(C) \subset Z^i(C)$, so the quotient is defined. For cycles z, z', write $[z] = z + B^i(C)$; then

$$[z] = [z'] \iff z - z' = d^{i-1}w \quad \text{for some } w \in C^{i-1},$$
$$H^i(C) = 0 \iff \ker d^i = \operatorname{im} d^{i-1}.$$

Thus cohomology measures failure of exactness at the specified degree.

For the complex $C^0 = \mathbb{Z}$, $C^1 = \mathbb{Z}$, $d^0(a) = 6a$, we obtain

$$H^0(C) = 0, \qquad H^1(C) = \mathbb{Z}/6\mathbb{Z}.$$

Changing the differential to zero gives $\mathbb{Z}$ in both degrees. The modules alone do not determine the cohomology; the maps are indispensable.

The notation $C^\bullet$ denotes the whole complex, with the dot standing for its varying degree. A *cochain map* $f : C^\bullet \to D^\bullet$ is a family $f^i : C^i \to D^i$ with $d_D^i f^i = f^{i+1} d_C^i$. It sends cycles to cycles because $d_D f(z) = f(d_C z) = 0$, and boundaries to boundaries because $f(d_C w) = d_D f(w)$. Therefore it induces $H^i(f)[z] = [f(z)]$. Identity and composition are preserved, making H^i a functor from complexes to modules.

Two cochain maps f, g are *homotopic* if there are maps $h^i : C^i \to D^{i-1}$ such that

$$f^i - g^i = d_D^{i-1} h^i + h^{i+1} d_C^i.$$

For a cycle z, their difference on z is $d_D h(z)$, a boundary. Consequently homotopic maps induce equal maps on cohomology. This is a precise reason why changing an auxiliary choice can leave cohomology unchanged.

A complete cohomology calculation with three nonzero terms

Consider the complex in degrees $0, 1, 2$

$$C^\bullet : \quad \mathbb{Z} \xrightarrow{d^0} \mathbb{Z}^2 \xrightarrow{d^1} \mathbb{Z}, \qquad d^0(n) = (2n, 0), \quad d^1(a, b) = 3b.$$

First verify $d^1 d^0(n) = 0$; without this identity cohomology is not defined. Next compute cycles and boundaries separately:

i	Z^i	B^i	H^i
0	0	0	0
1	$\mathbb{Z} \oplus 0$	$2\mathbb{Z} \oplus 0$	$\mathbb{Z}/2\mathbb{Z}$
2	$\mathbb{Z}$	$3\mathbb{Z}$	$\mathbb{Z}/3\mathbb{Z}$.

For $i = 2$ the outgoing differential is zero, so every element of $\mathbb{Z}$ is a cycle. For $i = 0$ the incoming differential has zero image. These endpoint conventions are part of the calculation.

The induced maps on cohomology also require representatives. Multiplication by an integer r in every degree is a cochain map, and it induces multiplication by r on both quotient groups. More generally a proposed family of matrices must commute with the differentials before it can be applied to classes of cycles.

Acyclic need not mean contractible

A complex is acyclic if all its cohomology groups vanish. It is contractible if its identity map is homotopic to zero. Contractibility implies acyclicity, since for a cycle z the identity $z = dh(z) + h(dz)$ expresses z as a boundary.

The converse is false. The exact complex

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

is acyclic. If it were contractible, the homotopy identity in the last nonzero degree would provide a section of $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. Such a section does not exist. Cohomology detects exactness, while a contracting homotopy supplies compatible splittings, which are additional data.

13.4 Exact sequences of complexes

A sequence of complexes is exact if it is exact in every degree. Kernels and cokernels of cochain maps are constructed degree by degree: the commuting-square condition makes the differentials restrict to kernels and descend to cokernels. The category of complexes of modules is therefore abelian. The next lecture explains why a short exact sequence of complexes usually produces a *long* sequence after taking cohomology. The extra arrows record the obstruction to lifting a cycle to a cycle.

13.5 Exercises

Exercise 13.1. Let $C^0 = \mathbb{Z}$, $C^1 = \mathbb{Z}^2$, $C^2 = \mathbb{Z}$, and $C^i = 0$ in all other degrees, with $d^0(a) = (2a, 0)$ and $d^1(b, c) = 3c$. Check the complex condition and compute all cohomology groups.

Exercise 13.2. Let R be a commutative ring and M an R-module. Show that

$$\mathrm{Hom}_R(R, M) \longrightarrow M, \qquad f \longmapsto f(1),$$

is a natural isomorphism. Identify its inverse.

Exercise 13.3. Let M be a module over a commutative ring R. Let C have $C^0 = C^1 = M$, differential id_M, and zero terms in every other degree. Construct a homotopy between id_C and the zero map. Why does this imply $H^i(C) = 0$?

Exercise 13.4 (Cohomology of a complex)**.** Let $C^0 = \mathbb{Z}^2$, $C^1 = \mathbb{Z}$, with $d^0(a, b) = 2a + 4b$, and all other terms zero. Compute all $H^i(C^\bullet)$. What changes if the differential is replaced by $d^0(a, b) = a + 2b$?

Lecture 14
The long exact sequence in cohomology

The connecting homomorphism will be constructed explicitly, with a proof of exactness at every term of the resulting sequence. The construction requires both the existence of suitable representatives and independence of their choices. Reading: [2], Chapter III, Section 1; here we give a complete proof for complexes of modules.

14.1 The connecting homomorphism

Suppose

$$0 \longrightarrow A^\bullet \xrightarrow{i} B^\bullet \xrightarrow{p} C^\bullet \longrightarrow 0$$

is a short exact sequence of cochain complexes. In each degree q, the map i^q is injective, p^q is surjective, and $\ker p^q = \operatorname{im} i^q$. The maps also commute with the differentials. We write d_A, d_B, d_C when it matters which complex is being used.

A cycle is also called a *closed element*. Starting from $c \in Z^q(C)$, lift it and apply the differential:

$$\begin{aligned} p(b) &= c && \text{(surjectivity of } p^q\text{)}, \\ p(d_B b) &= d_C c = 0, \\ d_B b &= i(a) && \text{for a unique } a \in A^{q+1}. \end{aligned}$$

The last line uses $\ker p^{q+1} = \operatorname{im} i^{q+1}$ and injectivity of i^{q+1}. The element a measures the failure of b to be closed.

Theorem 14.1 (Long exact cohomology sequence)**.** *A short exact sequence of complexes of R-modules gives an exact sequence*

$$\cdots \longrightarrow H^q(A) \xrightarrow{H^q(i)} H^q(B) \xrightarrow{H^q(p)} H^q(C) \xrightarrow{\delta^q} H^{q+1}(A) \longrightarrow H^{q+1}(B) \longrightarrow \cdots.$$

The connecting homomorphism is defined by $\delta^q[c] = [a]$, using the lifting procedure above.

Proof. We first justify the definition and then prove exactness at each of the three kinds of terms.

Step 1: the output is a cycle. Applying d_B to $i(a) = d_B b$ gives

$$i(d_A a) = d_B i(a) = d_B^2 b = 0.$$

Since i^{q+2} is injective, $d_A a = 0$. Therefore $[a]$ is a cohomology class.

Step 2: independence of the lift. If b' is another lift of c, then $p(b' - b) = 0$. Exactness gives $b' - b = i(u)$ for some $u \in A^q$. The corresponding outputs satisfy

$$i(a' - a) = d_B(b' - b) = i(d_A u),$$

so $a' - a = d_A u$. They differ by a boundary and represent the same class.

Step 3: independence of the cycle representative. Replace c by $c + d_C v$, where $v \in C^{q-1}$. Choose a lift $w \in B^{q-1}$ of v. Then $b + d_B w$ lifts $c + d_C v$, and $d_B(b + d_B w) = d_B b$. With this choice the output a is unchanged. Step 2 handles every other lift. Hence δ^q is well-defined on cohomology classes.

Step 4: linearity. If b_1, b_2 lift cycles c_1, c_2, then $rb_1 + sb_2$ lifts $rc_1 + sc_2$ and its differential is $i(ra_1 + sa_2)$. Thus δ^q is R-linear.

Step 5: exactness at $H^q(B)$. The composite $H^q(p)H^q(i)$ is zero because $pi = 0$. Conversely, suppose $b \in B^q$ is a cycle and $[p(b)] = 0$. There is $v \in C^{q-1}$ with $p(b) = d_C v$. Lift v to $w \in B^{q-1}$. Then

$$p(b - d_B w) = 0,$$

so $b - d_B w = i(a)$ for some $a \in A^q$. Its differential is zero, hence injectivity of i gives $d_A a = 0$. Finally $[b] = [i(a)]$. This proves that the kernel at $H^q(B)$ equals the preceding image.

Step 6: exactness at $H^q(C)$. If b is already a cycle, it can be used as the lift of $p(b)$, producing output zero. Therefore $\delta^q H^q(p) = 0$. Conversely, if $\delta^q[c] = 0$, choose b and a as in the construction. Vanishing means $a = d_A u$ for some $u \in A^q$. Then

$$d_B(b - i(u)) = i(a) - i(d_A u) = 0, \qquad p(b - i(u)) = c.$$

The adjusted lift is a cycle, so $[c]$ lies in the image of $H^q(p)$.

Step 7: exactness at $H^{q+1}(A)$. A connecting class $[a]$ maps to zero because $i(a) = d_B b$ is a boundary. Conversely, let $a \in A^{q+1}$ be a cycle whose image class is zero. Write $i(a) = d_B b$ for some $b \in B^q$. Then $c = p(b)$ is a cycle because $d_C c = p(d_B b) = p(i(a)) = 0$. The construction gives $\delta^q[c] = [a]$. These three checks, with q varying over all integers, prove exactness everywhere. □

This procedure is called a *diagram chase*. Its construction can be recorded as

$$[c] \;\longmapsto\; b \;\longmapsto\; d_B b = i(a) \;\longmapsto\; [a] = \delta^q[c], \qquad p(b) = c.$$

The intermediate lift b is a choice; the final class $[a]$ is independent of that choice.

14.2 A two-term complex over the integers

Fix an integer $n \geq 2$. Let B be the two-term complex

$$B^0 = \mathbb{Z} \xrightarrow{n} B^1 = \mathbb{Z}.$$

Let A have only $A^1 = \mathbb{Z}$, and let C have only $C^0 = \mathbb{Z}$. Define $i^1 = \mathrm{id}_{\mathbb{Z}}$ and $p^0 = \mathrm{id}_{\mathbb{Z}}$; other maps are zero. The sequence $0 \to A \to B \to C \to 0$ is exact in degree zero and degree one, and trivially exact elsewhere.

Its cohomology gives

$$0 \longrightarrow H^0(B) \longrightarrow \mathbb{Z} \xrightarrow{\delta^0} \mathbb{Z} \longrightarrow H^1(B) \longrightarrow 0.$$

For $c \in C^0 = \mathbb{Z}$,

$$b = c, \qquad d_B b = nc = i(nc), \qquad \delta^0(c) = nc.$$

The long exact sequence becomes

$$0 \longrightarrow 0 \longrightarrow \mathbb{Z} \xrightarrow{n} \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow 0.$$

The connecting arrow need not vanish even though the sequences of modules in individual degrees split. The splittings must commute with differentials to force it to vanish.

14.3 Reduction modulo an integer: a second connecting map

Let K be a complex of free abelian groups. Multiplication by n is injective in every degree, so

$$0 \longrightarrow K \xrightarrow{n} K \longrightarrow K/nK \longrightarrow 0$$

is a short exact sequence of complexes. A cycle modulo n is represented by an element $b \in K^q$ satisfying $db \in nK^{q+1}$. Write $db = na$. The connecting map is

$$\beta^q : H^q(K/nK) \longrightarrow H^{q+1}(K), \qquad [\bar{b}] \longmapsto [a].$$

This is often called a *Bockstein homomorphism*. The apparently informal instruction "lift, differentiate, and divide by n" is justified because the groups are free and the general proof above handles all choices.

For $K = (\mathbb{Z} \xrightarrow{n} \mathbb{Z})$ in degrees zero and one, the reduced differential is zero. Hence

$$H^0(K/nK) = \mathbb{Z}/n\mathbb{Z}, \qquad H^1(K) = \mathbb{Z}/n\mathbb{Z}.$$

A lift b has differential nb, so dividing by n gives b. Thus β^0 is the identity between these two copies of $\mathbb{Z}/n\mathbb{Z}$. It detects torsion that would be invisible if one looked only at the reduced differential.

The connecting map as a precise obstruction

Write a short exact sequence of complexes as

$$0 \to A^\bullet \xrightarrow{i} B^\bullet \xrightarrow{q} C^\bullet \to 0.$$

For a cycle $c \in C^n$, lift it to $b \in B^n$. Then $q(db) = dc = 0$, so $db = i(a)$ for a unique $a \in A^{n+1}$. Injectivity of i and $d^2 b = 0$ imply $da = 0$. The class $[a]$ is $\delta[c]$.

Suppose $\delta[c] = 0$. Then $a = du$ for some $u \in A^n$, and

$$d(b - i(u)) = db - i(du) = 0, \qquad q(b - i(u)) = c.$$

Thus the chosen cycle c itself admits a cycle lift after correcting its arbitrary lift by an element of A^n. Conversely, if c has a cycle lift b', then the construction with b' gives $\delta[c] = 0$. This explains the exactness identity

$$\ker\bigl(\delta : H^n(C) \to H^{n+1}(A)\bigr) = \operatorname{im}\bigl(H^n(B) \to H^n(C)\bigr).$$

If a representative of $[c]$ is changed by a boundary, lift that boundary as well; the same conclusion holds for its class. There is no requirement that the original arbitrary lift b already be a cycle.

A connecting map computed by division

Let $C^\bullet$ be $\mathbb{Z} \xrightarrow{\times m} \mathbb{Z}$ in degrees $0, 1$, and let $r > 0$. Multiplication by r is degreewise injective and gives

$$0 \to C^\bullet \xrightarrow{\times r} C^\bullet \to C^\bullet / rC^\bullet \to 0.$$

A degree-zero cycle in the quotient is a class $\bar{a} \in \mathbb{Z}/r\mathbb{Z}$ satisfying $r \mid ma$. Lift it to $a \in \mathbb{Z}$. Its differential is $ma = r(ma/r)$, so the boundary map is

$$\delta : \ker(\times m : \mathbb{Z}/r\mathbb{Z} \to \mathbb{Z}/r\mathbb{Z}) \to \mathbb{Z}/m\mathbb{Z}, \qquad \bar{a} \longmapsto \overline{ma/r}.$$

Changing the lift to $a + rb$ changes the answer by mb, which vanishes in $\mathbb{Z}/m\mathbb{Z}$. For $m = 6$, $r = 4$, the source is $\{\bar{0}, \bar{2}\}$, and $\delta(\bar{2}) = \bar{3} \in \mathbb{Z}/6\mathbb{Z}$. Its image is exactly the kernel of multiplication by four on $\mathbb{Z}/6\mathbb{Z}$, as the long exact sequence requires.

14.4 Naturality of the connecting homomorphism

A map between short exact sequences is a commutative diagram of cochain maps

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{i} & B & \xrightarrow{p} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle w} & & \\ 0 & \longrightarrow & A' & \xrightarrow[i']{} & B' & \xrightarrow[p']{} & C' & \longrightarrow & 0. \end{array}$$

The connecting maps satisfy

$$H^{q+1}(u)\delta^q = \delta'^q H^q(w).$$

Indeed, from $p(b) = c$ and $d_B b = i(a)$ we obtain

$$\begin{aligned} p'v(b) &= w(p(b)) = w(c), \\ d_{B'}v(b) &= v(d_B b) = v(i(a)) = i'(u(a)), \\ \delta'^q[w(c)] &= [u(a)] = H^{q+1}(u)\delta^q[c]. \end{aligned}$$

This compatibility is the naturality of the long exact sequence.

The theorem also holds in every abelian category. Its formulation uses kernels, cokernels, and images, so it makes sense there. Our proof was explicitly a proof for modules, where elements and lifts are available. The general categorical theorem requires a categorical diagram argument or an appropriate embedding theorem; that extension is not needed for our present computations. A further application will be given after the geometric cohomology theory of Lecture 26 has been defined.

14.5 Exercises

Exercise 14.1. Let $0 \to A \xrightarrow{i} B \xrightarrow{p} C \to 0$ be a short exact sequence of cochain complexes. Suppose it has a *section*, meaning a cochain map $s : C \to B$ with $ps = \mathrm{id}_C$. Prove that all connecting homomorphisms vanish and that $H^q(B) \cong H^q(A) \oplus H^q(C)$.

Exercise 14.2. Let $K = (\mathbb{Z} \xrightarrow{6} \mathbb{Z})$ be concentrated in degrees 0 and 1, and reduce it modulo 4. For the short exact sequence $0 \to K \xrightarrow{4} K \to K/4K \to 0$, compute the connecting map $\beta^0 : H^0(K/4K) \to H^1(K)$.

Exercise 14.3. Let $0 \to A \xrightarrow{i} B \xrightarrow{p} C \to 0$ be a short exact sequence of cochain complexes. If $H^q(A) = H^{q+1}(A) = 0$, prove that $H^q(B) \to H^q(C)$ is an isomorphism. If every cohomology group of B is zero, identify $H^q(C)$.

Lecture 15
A first course in topology and manifolds

This lecture introduces the topology required to discuss solution sets of polynomial equations and their local pieces. No previous topology is assumed. The manifold examples provide a comparison for the algebraic charts introduced later. Reading: the topological definitions accompanying [2], Chapter I, Sections 1–3; the manifold discussion here is elementary background.

15.1 Topological spaces and continuous maps

On $\mathbb{R}$, an interval (a,b) allows a point to move slightly without leaving the interval. In $\mathbb{R}^n$, put

$$B(x,\varepsilon) = \{y : \|y-x\| < \varepsilon\}, \qquad \|y-x\| = \sqrt{\sum_{j=1}^{n} (y_j - x_j)^2}.$$

Then $U \subset \mathbb{R}^n$ is Euclidean open if

$$\forall x \in U \ \exists \varepsilon > 0 : \quad B(x,\varepsilon) \subset U.$$

The formal properties of open sets can be stated without a distance.

Definition 15.1. A *topology* on a set X is a collection $\mathscr{T}$ of subsets, called open sets, satisfying

$$\varnothing, X \in \mathscr{T}, \qquad \bigcup_{\lambda \in \Lambda} U_\lambda \in \mathscr{T} \quad (U_\lambda \in \mathscr{T}),$$
$$U_1 \cap \cdots \cap U_r \in \mathscr{T} \quad (U_1, \ldots, U_r \in \mathscr{T}).$$

The union may have any index set Λ; the intersection is finite. A set F is *closed* if $X \setminus F \in \mathscr{T}$. A *neighborhood* N of x contains an open set U with $x \in U \subset N$.

Closed sets are closed under arbitrary intersections and finite unions. Infinite intersections of open sets need not be open: in $\mathbb{R}$,

$$\bigcap_{n\geq 1}(-1/n,1/n)=\{0\}.$$

The singleton is not open because every interval around zero contains nonzero points. A set may be both open and closed: $\varnothing$ and X always are. A set may also be neither; $[0,1)$ in $\mathbb{R}$ is an example.

The *discrete topology* declares every subset open. The *indiscrete topology* has only $\varnothing$ and X open. These examples warn us that topology is extra structure on a set. The same underlying set may carry very different topologies.

A family $\mathscr{B}$ of open sets is a *basis* if

$$U\in\mathscr{T} \quad\Longrightarrow\quad U=\bigcup_{\substack{B\in\mathscr{B}\\ B\subset U}} B.$$

Equivalently, $x\in U\in\mathscr{T}$ implies $x\in B\subset U$ for some $B\in\mathscr{B}$. Open balls form a basis for Euclidean space. A property preserved by arbitrary unions can therefore be checked on basis elements.

Definition 15.2. The *subspace topology* on $Y\subset X$ has open sets $Y\cap U$ for U open in X. A map $f:X\to Y$ is *continuous* if $f^{-1}(V)$ is open whenever V is open in Y. A *homeomorphism* is a bijective continuous map with continuous inverse.

The inverse-image formulation explains why continuous maps compose: $(g\circ f)^{-1}(W)=f^{-1}(g^{-1}(W))$. It is equivalent to requiring inverse images of closed sets to be closed. Topological spaces and continuous maps form a category, denoted Top.

A bijective continuous map need not be a homeomorphism. On a set with at least two elements, the identity from the discrete topology to the indiscrete topology is continuous, but its inverse is not. The word "inverse" in the definition of a homeomorphism therefore adds a real requirement.

15.2 Closure, density, connectedness, and separation

Definition 15.3. The *closure* $\overline{S}$ of $S\subset X$ is the intersection of all closed subsets containing S. The set S is *dense* if $\overline{S}=X$.

The useful tests for closure and density are

$$x\in\overline{S}\iff U\cap S\neq\varnothing \quad\text{for every open } U\ni x,$$
$$\overline{S}=X\iff U\cap S\neq\varnothing \quad\text{for every nonempty open } U.$$

Indeed, an open neighborhood U missing S has closed complement containing S and missing x. Conversely, if $x\notin\overline{S}$, take $U=X\setminus\overline{S}$. For example,

$$a<b\implies(a,b)\cap\mathbb{Q}\neq\varnothing\implies\overline{\mathbb{Q}}=\mathbb{R}.$$

A space is *Hausdorff* if distinct points have disjoint open neighborhoods. Euclidean spaces are Hausdorff: balls of radius less than one third of the distance between two points are disjoint. In a Hausdorff space each singleton is closed. Indeed, for every $y \neq x$ choose an open neighborhood of y avoiding x, and take their union to obtain $X \setminus \{x\}$.

A space is *connected* if it cannot be written as a union of two disjoint nonempty open subsets. Equivalently, it has no nonempty proper subset that is both open and closed. A *path* from x to y is a continuous map $\gamma : [0,1] \to X$ with the indicated endpoint values. A space is *path connected* if such a path exists for every pair of points.

Path connectedness implies connectedness. To see this, a decomposition into two disjoint nonempty open sets would pull back along a path between them to a separation of $[0,1]$. An interval has no such separation, as follows from the least-upper-bound property of $\mathbb{R}$ or the intermediate value theorem. We will not need the converse, which is false in general.

15.3 Covers, compactness, and products

An *open cover* of X is a family of open sets whose union is X. A space is *quasi-compact* if every open cover has a finite subcover. Some authors use "compact" for this definition; in these notes "compact" will mean quasi-compact and Hausdorff when discussing general spaces. This distinction matters because spaces arising from polynomial equations will usually fail to be Hausdorff.

The closed interval $[0,1]$ is compact by the Heine–Borel theorem from real analysis. The open interval $(0,1)$ is not quasi-compact: the increasing open sets $(1/n, 1)$, for $n \geq 2$, cover it, but no finite selection covers points sufficiently close to zero. Compactness is a property of all open covers, not a statement that a drawing looks bounded. In particular, it must not be confused with having finitely many points.

The *product topology* on $X \times Y$ has basis $U \times V$ with U, V open. The projections π_X, π_Y are continuous because

$$\pi_X^{-1}(U) = U \times Y, \qquad \pi_Y^{-1}(V) = X \times V.$$

For $h = (h_X, h_Y) : T \to X \times Y$,

$$h^{-1}(U \times V) = h_X^{-1}(U) \cap h_Y^{-1}(V).$$

Thus h is continuous exactly when both coordinate maps are continuous: use this formula in one direction and $h_X = \pi_X h$, $h_Y = \pi_Y h$ in the other.

The *quotient topology* associated with a surjective set map $q : X \to Y$ declares $V \subset Y$ open exactly when $q^{-1}(V)$ is open. This is useful when making a space by identifying points. A map $f : Y \to Z$ is continuous exactly when $f \circ q$ is continuous, directly from the inverse-image definition.

A useful theorem about continuous bijections

Suppose X is quasi-compact, Y is Hausdorff, and $f : X \to Y$ is a continuous bijection. Then f is a homeomorphism. We give the intermediate facts, because the two topological hypotheses do different jobs.

First, a closed subset $C \subseteq X$ is quasi-compact. Given an open cover of C in its subspace topology, write its members as $C \cap U_i$ and add $X \setminus C$ to obtain a cover of X. A finite subcover of X leaves a finite subcover of C. Second, a continuous image of a quasi-compact space is quasi-compact: pull an open cover back to the domain and then push the finite selection forward.

Third, a quasi-compact subset K of a Hausdorff space is closed. For $y \notin K$ and each $x \in K$, choose disjoint neighborhoods $U_x \ni x$ and $V_x \ni y$. Finitely many $U_{x_1}, \ldots, U_{x_r}$ cover K. Then $\bigcap_i V_{x_i}$ is a neighborhood of y disjoint from K. Hence the complement of K is open.

Apply these facts to a closed $C \subseteq X$. The set $f(C)$ is quasi-compact and therefore closed in Y. Thus f is a closed bijection. For its inverse g, the preimage $g^{-1}(C) = f(C)$ is closed whenever C is closed, so g is continuous. This proves the theorem. It cannot be applied to a general Zariski space by ignoring the Hausdorff hypothesis.

Noetherian topology and finite covers of open subsets

A space is called Noetherian if every descending chain of closed subsets stabilizes, or equivalently every ascending chain of open subsets stabilizes. Every open subset of a Noetherian space is quasi-compact.

To prove this, cover an open U by open subsets U_i of U. Because U is itself open in X, each U_i is open in X. Among finite unions of the U_i, choose one maximal under inclusion. Such a maximal union exists: otherwise repeatedly choosing a strictly larger finite union would produce a nonstabilizing ascending chain. If that finite union omitted a point of U, adding a member of the cover containing the point would enlarge it, a contradiction. Thus a finite subcollection covers U.

Conversely, if every open subset is quasi-compact and $U_1 \subseteq U_2 \subseteq \cdots$, the union U is covered by this chain. A finite subcover has a largest member, say U_N, so $U = U_N$ and the chain stabilizes. This proves the equivalence. A Noetherian topological space may have infinitely many points; the condition concerns chains of subsets, not their cardinalities.

15.4 Manifolds and coordinate charts

Definition 15.4. An n-dimensional *topological manifold* is a Hausdorff space with a countable basis in which every point has an open neighborhood homeomorphic to an open subset of $\mathbb{R}^n$. Such a homeomorphism $\phi : U \to V \subset \mathbb{R}^n$ is a *chart*. A family of charts covering the space is an *atlas*.

The countable-basis condition excludes unnecessarily large examples and is satisfied by the spaces encountered in elementary geometry. It is not an algebraic finiteness condition such as being Noetherian.

On an overlap of two charts, a point has two coordinate descriptions. The *transition map* is

$$\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \longrightarrow \phi_j(U_i \cap U_j).$$

Its domain and target are open subsets of Euclidean space. A function on a Euclidean open set is *infinitely differentiable*, or C^∞, if all its partial derivatives of every finite order exist and are continuous. A vector-valued map has this property when each coordinate function does. A *smooth manifold* has an atlas for which transition maps are infinitely differentiable; charts smoothly compatible with that atlas are understood to be included. A map between smooth manifolds is smooth if its expressions in local coordinates are infinitely differentiable. Transition compatibility makes this independent of which compatible charts are used.

For $S^1 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, four charts suffice:

open subset	coordinate	inverse on $(-1,1)$
$y > 0$	x	$(t, \sqrt{1-t^2})$
$y < 0$	x	$(t, -\sqrt{1-t^2})$
$x > 0$	y	$(\sqrt{1-t^2}, t)$
$x < 0$	y	$(-\sqrt{1-t^2}, t)$

They cover the circle. On each overlap, transition functions are restrictions of $t \mapsto \pm\sqrt{1-t^2}$ or the identity, hence are smooth. Every chart resembles an interval; several charts describe the whole circle.

The same method will be used for polynomial equations: cover a solution set by open pieces, describe the functions on each piece, and check agreement on overlaps. The pieces need not be open subsets of Euclidean space. Lecture 16 defines their Zariski topology, and Lecture 21 gives an algebraic criterion for the exceptional points at which local equations fail to impose the expected number of independent linear conditions.

15.5 Exercises

Exercise 15.1. Give the subspace topology on $X = \{0, 1\} \subset \mathbb{R}$. Is X connected? Is it compact?

Exercise 15.2. Show that the continuous image of a quasi-compact space is quasi-compact, and that the continuous image of a connected space is connected.

Exercise 15.3. Let X be infinite and give it the *cofinite topology*: the closed sets are X and all finite subsets. Show that it is a topology, is quasi-compact, and is not Hausdorff.

Lecture 16
The Zariski topology and the spectrum of a ring

Polynomial equations and prime ideals each define a topology. We will distinguish the classical points of an algebraic set from the additional points in a spectrum. Reading: [2], Chapter I, Section 1, and Chapter II, Section 2; [1], Chapter 1, for spectra and prime ideals.

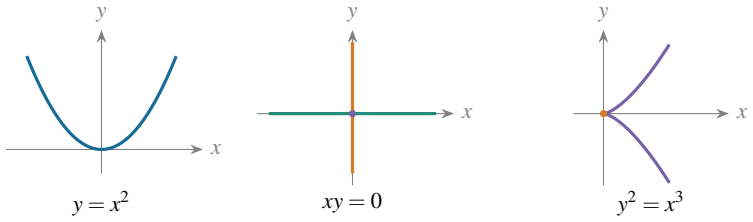


Fig. 16.1 Real points of a parabola, $V(xy) = V(x) \cup V(y)$, and a cusp. Marked origins are the crossing point and the cusp point.

16.1 Algebraic sets and the Zariski topology

Fix an algebraically closed field k, meaning that every nonconstant polynomial in one variable over k has a root in k. Write $\mathbb{A}^n(k) = k^n$, often abbreviated $\mathbb{A}^n$. For an ideal $I \subset k[x_1, \ldots, x_n]$, define

$$V(I) = \{a \in k^n : f(a) = 0 \text{ for every } f \in I\}.$$

If $I = (f_1, \ldots, f_r)$, then

$$V(I) = \{a : f_1(a) = \cdots = f_r(a) = 0\},$$

because $f = \sum_i h_i f_i$ gives $f(a) = \sum_i h_i(a) f_i(a)$. Such a zero set is an *affine algebraic set*. By the Hilbert basis theorem, every ideal admits finitely many generators, so finitely many equations suffice.

Proposition 16.1. *The algebraic sets are the closed sets of a topology. More precisely,*

$$V(0) = \mathbb{A}^n, \quad V(1) = \varnothing, \quad \bigcap_\lambda V(I_\lambda) = V\Big(\sum_\lambda I_\lambda\Big), \quad V(I) \cup V(J) = V(IJ).$$

Proof. Evaluation of 0 and 1 gives the first two assertions. For intersections,

$$\begin{aligned} a \in \bigcap_\lambda V(I_\lambda) &\iff f(a) = 0 \quad (\lambda \in \Lambda,\ f \in I_\lambda) \\ &\iff h(a) = 0 \quad \big(h \in \textstyle\sum_\lambda I_\lambda\big). \end{aligned}$$

If $a \in V(I) \cup V(J)$, every fg with $f \in I$, $g \in J$ vanishes at a, hence so does every element of IJ. Conversely,

$$\begin{aligned} a \notin V(I) \cup V(J) &\implies \exists f \in I,\ g \in J : f(a)g(a) \neq 0 \\ &\implies a \notin V(IJ). \end{aligned}$$

These identities are the closed-set axioms. □

This is the *Zariski topology*. A *principal open set* is $D(f) = \mathbb{A}^n \setminus V(f)$, the locus where f is nonzero. Such opens form a basis, because

$$\mathbb{A}^n \setminus V(I) = \bigcup_{f \in I} D(f).$$

Moreover $D(f) \cap D(g) = D(fg)$. On an algebraic subset X, we use the subspace topology and write $D_X(f) = X \cap D(f)$.

On $\mathbb{A}^1$, every nonzero polynomial has finitely many roots and every finite subset is the zero set of a product of linear factors. Thus the Zariski topology is cofinite. The field k is infinite: if a field had finitely many elements $a_1, \dots, a_r$, the polynomial $\prod_i (t - a_i) - 1$ would have no root and the field could not be algebraically closed. Consequently nonempty opens in $\mathbb{A}^1$ always intersect. This topology is far coarser than the Euclidean topology on $\mathbb{C}$.

For any subset $Y \subset \mathbb{A}^n$, its *vanishing ideal* is

$$I(Y) = \{ f \in k[x_1, \dots, x_n] : f(a) = 0 \text{ for every } a \in Y \}.$$

It is an ideal because sums and polynomial multiples preserve vanishing. If f^r vanishes on Y, then f vanishes there, since its values lie in a field. Thus $I(Y)$ is radical. For an algebraic set Y, the quotient $k[Y] = k[x_1, \dots, x_n]/I(Y)$ is its *coordinate ring*: its elements are the polynomial functions on Y. It is reduced, since radicality of $I(Y)$ excludes nonzero nilpotents in the quotient. These definitions do not require the Nullstellensatz; that theorem will later identify $I(V(J))$ for an arbitrary ideal J.

16.2 Irreducible spaces and components

Definition 16.2. A nonempty topological space X is *irreducible* if it is not the union of two proper closed subsets. An affine *variety* in these notes is a nonempty irreducible affine algebraic set, equipped with its reduced coordinate ring. An algebraic set is allowed to be reducible.

Standing conventions for the geometric lectures. From this lecture onward the ground field k is algebraically closed unless a different field is explicitly specified. An algebraic set may be reducible. The word *variety* always means a nonempty irreducible algebraic set with its reduced coordinate rings; reducible examples and rings with nilpotents will be identified explicitly. Later arithmetic examples over $\mathbb{Q}$ or finite fields will state their ground fields separately.

Taking complements gives the equivalent tests

$$\begin{aligned} X \text{ irreducible} &\iff U \cap V \neq \varnothing \quad \text{for nonempty open } U, V \\ &\iff \overline{U} = X \quad \text{for every nonempty open } U. \end{aligned}$$

For the second equivalence, U is not dense exactly when $X \setminus \overline{U}$ is a nonempty open set disjoint from U.

Every nonempty open subset U of an irreducible space is itself irreducible. Indeed, two nonempty opens in U are also open in X, because U is open, so they intersect. In particular, prescribing a function on a nonempty open part of a variety prescribes it on a dense subset.

Irreducibility implies connectedness: a separation into two nonempty disjoint open sets would also express the space as the union of two proper closed sets. The reverse implication fails. The algebraic set

$$X = V(xy) \subset \mathbb{A}^2$$

is the union of its two coordinate axes, so it is reducible. Each axis is an irreducible affine line and hence connected. Their intersection contains the origin, so their union is connected. The last assertion follows because any separation of the union would have to place each connected axis entirely on one side, and their common point forces the same side for both.

An *irreducible component* of a space is a maximal irreducible closed subset, where maximal means maximal under inclusion. A topological space is *Noetherian* if every descending chain of closed subsets becomes constant.

Proposition 16.3. *Every algebraic set has finitely many irreducible components: it is a finite union of irreducible closed subsets, uniquely determined after omitting any member contained in another.*

Proof. A descending chain of algebraic closed sets stabilizes. Indeed, their vanishing ideals form an ascending chain in a Noetherian polynomial ring, and any algebraic set Y equals $V(I(Y))$: if $Y = V(J)$, then $J \subset I(Y)$ gives one inclusion and the other is immediate. Hence the underlying topological space is Noetherian.

If some closed set were not a finite union of irreducible closed subsets, choose a minimal such set by the descending-chain condition. It is reducible, so it is a union of two proper closed subsets. Each smaller subset is a finite union of irreducible ones by minimality, a

contradiction. For uniqueness, if $Y = \bigcup_i Y_i = \bigcup_j Z_j$, then $Y_i = \bigcup_j (Y_i \cap Z_j)$. Irreducibility implies $Y_i \subset Z_j$ for some j, using induction for finite unions. Similarly $Z_j \subset Y_\ell$ for some ℓ. With no containments among the Y_i, we have $\ell = i$ and equality throughout. The two collections coincide. □

The affine line is not Hausdorff

Assume k is an infinite algebraically closed field. A proper closed subset of $\mathbb{A}^1_k$ is finite: if its ideal contains a nonzero polynomial, that polynomial has only finitely many roots. Conversely every finite set $\{a_1, \ldots, a_r\}$ is the zero set of $\prod_i (t - a_i)$. Thus the nonempty opens are precisely the complements of finite subsets.

Two nonempty opens must intersect, since the union of their finite complements cannot exhaust the infinite field. Consequently two distinct points cannot have disjoint neighborhoods. The line is not Hausdorff, although each point is closed. This shows why the condition "all points are closed" is weaker than Hausdorffness.

The same calculation proves that every infinite subset of $\mathbb{A}^1_k$ is dense. A sequence of pairwise distinct points even converges, in the topological sense, to every point: every cofinite neighborhood contains all but finitely many sequence terms. There is no contradiction with uniqueness of limits in familiar analysis, because that uniqueness uses Hausdorffness. Sequences are therefore a poor substitute for equations when studying Zariski closure.

Connected and irreducible are different

An irreducible space is connected. Indeed, two disjoint nonempty opens whose union is the space would have disjoint nonempty complements, expressing the space as a union of two proper closed subsets. More directly, in an irreducible space every two nonempty open subsets intersect.

The union of the axes $V(xy) \subseteq \mathbb{A}^2_k$ is reducible, since it is $V(x) \cup V(y)$. Nevertheless it is connected. Each axis is an irreducible affine line, hence connected, and they share the origin. A separation of their union would have to place each connected axis entirely in one of the two sides. Their common point forces them into the same side, leaving the other empty. Irreducible components may therefore meet; connected components are coarser pieces.

16.3 The prime spectrum

For any commutative ring R, let $\operatorname{Spec} R$ be its set of prime ideals; prime ideals are proper, so $\operatorname{Spec} 0 = \varnothing$. Define

$$V_R(I) = \{\mathfrak{p} \in \operatorname{Spec} R : I \subset \mathfrak{p}\}, \qquad D_R(f) = \{\mathfrak{p} : f \notin \mathfrak{p}\}.$$

The same identities as above hold, using the defining implication $fg \in \mathfrak{p} \Rightarrow f \in \mathfrak{p}$ or $g \in \mathfrak{p}$. They define the Zariski topology on $\operatorname{Spec} R$, with basis $D_R(f)$. This construction currently specifies a topological space. The rings of functions on its open subsets are further data, constructed systematically in Lecture 25. The topology alone does not recover those rings.

Proposition 16.4. *The closure of a point $\mathfrak{p} \in \operatorname{Spec} R$ is $V_R(\mathfrak{p})$. Consequently $\mathfrak{p}$ is closed exactly when it is maximal.*

Proof. Since $V_R(\mathfrak{p})$ is closed and contains $\mathfrak{p}$, it suffices to observe

$$\mathfrak{p} \in V_R(I) \iff I \subset \mathfrak{p} \implies V_R(\mathfrak{p}) \subset V_R(I).$$

Consequently

$$\begin{aligned} \{\mathfrak{p}\} \text{ closed} &\iff V_R(\mathfrak{p}) = \{\mathfrak{p}\} \\ &\iff \text{no prime strictly contains } \mathfrak{p} \\ &\iff \mathfrak{p} \text{ maximal.} \end{aligned}$$

The last step uses the existence of a maximal ideal containing each proper ideal. □

For an integral domain, the prime (0) has closure all of $\operatorname{Spec} R$. Such a point is called a *generic point*: one point whose closure is the entire space. The generic point is not an ordinary coordinate tuple. For $R = k[t]$, the primes are (0) and $(t-a)$ for $a \in k$. The maximal ideals describe classical points, whereas (0) describes the whole irreducible line through its closure.

For $R = \mathbb{Z}$, the points are prime ideals, not integers on a number line:

$$\begin{aligned} \operatorname{Spec}\mathbb{Z} &= \{(0)\} \cup \{(p) : p \text{ a prime integer}\}, \\ D_{\mathbb{Z}}(6) &= \operatorname{Spec}\mathbb{Z} \setminus \{(2),(3)\}, \\ \overline{\{(0)\}} &= \operatorname{Spec}\mathbb{Z}, \qquad \overline{\{(p)\}} = \{(p)\}. \end{aligned}$$

The residue fields are $\kappa((0)) = \mathbb{Q}$ and $\kappa((p)) = \mathbb{F}_p$. In general,

$$\kappa(\mathfrak{p}) = R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}} = \operatorname{Frac}(R/\mathfrak{p}).$$

The last equality uses that $R/\mathfrak{p}$ is a domain and localization inverts precisely its nonzero elements.

Proposition 16.5. *Every spectrum is quasi-compact. A ring homomorphism $\varphi : R \to S$ induces a continuous map*

$$\operatorname{Spec} S \longrightarrow \operatorname{Spec} R, \qquad \mathfrak{q} \longmapsto \varphi^{-1}(\mathfrak{q}).$$

Proof. Refine the cover by principal opens. Then

$$\operatorname{Spec} R = \bigcup_{\lambda} D_R(f_\lambda) \implies (f_\lambda : \lambda \in \Lambda) = R$$
$$\implies 1 = \sum_{j=1}^{r} a_j f_{\lambda_j}$$
$$\implies \operatorname{Spec} R = \bigcup_{j=1}^{r} D_R(f_{\lambda_j}).$$

For the first implication, a proper ideal would lie in a maximal ideal missed by the cover. A prime containing all f_{λ_j} would contain 1, which proves the last implication. Each selected principal open lies in an original cover member, giving a finite subcover of the original cover.

The induced map $\varphi^*(\mathfrak{q}) = \varphi^{-1}(\mathfrak{q})$ is well-defined: $1 \notin \varphi^{-1}(\mathfrak{q})$, and

$$ab \in \varphi^{-1}(\mathfrak{q}) \implies \varphi(a)\varphi(b) \in \mathfrak{q} \implies a \text{ or } b \in \varphi^{-1}(\mathfrak{q}).$$

Finally $(\varphi^*)^{-1}(D_R(f)) = D_S(\varphi(f))$, so it is continuous. □

Generic points and the direction of specialization

For $\mathfrak{p} \in \operatorname{Spec} A$, a closed set $V(I)$ contains $\mathfrak{p}$ precisely when $I \subseteq \mathfrak{p}$. Therefore

$$\overline{\{\mathfrak{p}\}} = V(\mathfrak{p}) = \{\mathfrak{q} \in \operatorname{Spec} A : \mathfrak{p} \subseteq \mathfrak{q}\}.$$

One says that $\mathfrak{q}$ is a specialization of $\mathfrak{p}$ when $\mathfrak{q}$ lies in this closure. Bigger prime ideals describe more specialized points. In a domain, (0) has closure all of $\operatorname{Spec} A$ and is the generic point. In $k[x,y]$, the point (x) is generic for the line $x = 0$, and $(x, y-a)$ is a closed specialization on that line.

Every nonempty irreducible closed subset of a spectrum has one generic point. Write the subset as $V(I)$. Irreducibility is equivalent to $\sqrt{I}$ being prime: if $ab \in \sqrt{I}$, then

$$V(I) \subseteq V(ab) = V(a) \cup V(b).$$

Irreducibility implies containment in one of the two closed sets, so $a \in \sqrt{I}$ or $b \in \sqrt{I}$ by the prime-intersection description of radicals. Conversely, if $\mathfrak{p}$ is prime and $V(\mathfrak{p})$ were the union of two proper closed subsets, choose an equation from each that fails to vanish identically on $V(\mathfrak{p})$. Their product would belong to $\mathfrak{p}$, contradicting primality. The point $\sqrt{I}$ is then generic, and uniqueness follows because $V(\mathfrak{p}) = V(\mathfrak{q})$ forces $\mathfrak{p} = \mathfrak{q}$ for prime ideals.

Ring maps induce continuous maps in the opposite direction

For a unital ring map $\varphi : A \to B$, define

$$\varphi^* : \operatorname{Spec} B \to \operatorname{Spec} A, \qquad \mathfrak{q} \longmapsto \varphi^{-1}(\mathfrak{q}).$$

The contraction is proper since $\varphi(1) = 1 \notin \mathfrak{q}$. It is prime because ab in the contraction means $\varphi(a)\varphi(b) \in \mathfrak{q}$. Continuity follows from

$$(\varphi^*)^{-1}(D_A(f)) = D_B(\varphi(f)).$$

Contraction of ideals also shows $(\psi\varphi)^* = \varphi^*\psi^*$. This is the first contravariant part of the algebra–geometry correspondence. For a quotient $A \to A/I$ the image is $V(I)$; for a localization $A \to A_f$ the image is $D(f)$. Their descriptions follow from the two prime-correspondence theorems, and the basic-open formula proves that the corresponding identifications are homeomorphisms.

16.4 Exercises

Exercise 16.1. Describe $\operatorname{Spec}(k \times k)$, including its topology. Is it irreducible?

Exercise 16.2. Show that $\operatorname{Spec} k[\varepsilon]/(\varepsilon^2)$ has a single point. Compare it with $\operatorname{Spec} k$.

Exercise 16.3. Let k be algebraically closed. On the classical affine plane $\mathbb{A}^2_k$, show that the Zariski topology is different from the product of the Zariski topologies on two affine lines.

Lecture 17
The Nullstellensatz and affine algebraic sets

The Nullstellensatz gives a precise correspondence between polynomial equations and their solution sets. The central theorem explains why the correct algebraic object is a radical ideal, and why maximal ideals describe points over an algebraically closed field. Reading: [2], Chapter I, Section 1; [1], Chapter 5, for integrality and the Nullstellensatz.

17.1 Vanishing ideals and zero sets

Throughout this lecture, k is algebraically closed and $P = k[x_1, \ldots, x_n]$. For any subset $S \subset k^n$, put

$$I(S) = \{f \in P : f(a) = 0 \text{ for every } a \in S\}.$$

This is an ideal: for $f, g \in I(S)$ and $h \in P$,

$$(f+g)(a) = 0, \qquad (hf)(a) = 0 \qquad (a \in S).$$

The two operations reverse inclusions:

$$S \subset T \implies I(T) \subset I(S), \qquad J \subset K \implies V(K) \subset V(J).$$

Since k is a field,

$$f^m \in I(S) \implies f(a)^m = 0 \ (a \in S) \implies f \in I(S).$$

Thus $I(S)$ is radical. Recall that

$$\sqrt{J} = \{f : f^m \in J \text{ for some } m \geq 1\}.$$

It follows immediately that $\sqrt{J} \subset I(V(J))$. The difficult inclusion says that a function vanishing on all common zeros is already forced by the equations, after taking a power. We first need to understand maximal ideals.

17.2 Zariski's lemma

A field extension L/k is *algebraic* if each element satisfies a nonzero polynomial over k. A tuple is *algebraically independent* if no nonzero polynomial over k vanishes on it. Recall also that an element is integral over a ring if it satisfies a monic polynomial over that ring; the monic condition matters for rings that are not fields.

Lemma 17.1 (Zariski's lemma)**.** *If a field L is finitely generated as a k-algebra, then L is a finite-dimensional vector space over k. This assertion holds for any field k.*

Proof. Write $L = k[a_1, \dots, a_m]$. Choose a maximal algebraically independent subset of these generators, and relabel it $t_1, \dots, t_r$. Denote the remaining generators by $b_1, \dots, b_s$. Maximality means each b_j is algebraic over $K = k(t_1, \dots, t_r)$: a polynomial dependence on $t_1, \dots, t_r, b_j$ gives a genuine nonzero polynomial in b_j over K, because independence prevents all its coefficients from vanishing.

Suppose $r \geq 1$. Each b_j satisfies a monic equation with coefficients in K. There are only finitely many coefficients in the finitely many equations, so choose one nonzero polynomial $q \in k[t_1, \dots, t_r]$ divisible by all their denominators. Put

$$A = k[t_1, \dots, t_r, 1/q].$$

Now every b_j is integral over A. The ring $B = A[b_1, \dots, b_s]$ is a finite A-module: powers of each b_j reduce using its monic equation, leaving finitely many products of bounded powers as generators. Since L is a field, $1/q$ belongs to L. Thus

$$L = k[t_1, \dots, t_r, b_1, \dots, b_s] \subset B \subset L,$$

so $B = L$. In particular L is integral over A. Here we use the elementary finite-module criterion from Lecture 10: multiplication by an element of a finite algebra satisfies a monic equation, by the determinant argument on a finite generating set.

If a field is integral over a subring, that subring is a field. Indeed, for $0 \neq a \in A$, integrality of $a^{-1} \in L$ gives

$$(a^{-1})^d + c_{d-1}(a^{-1})^{d-1} + \cdots + c_0 = 0, \qquad c_j \in A.$$

Multiplying by a^{d-1} gives

$$a^{-1} = -c_{d-1} - c_{d-2}a - \cdots - c_0 a^{d-1} \in A.$$

We now show that our A cannot be a field when $r \geq 1$. There are infinitely many monic irreducibles in $k[t_1]$: if a finite list $p_1, \dots, p_N$ exhausted them, the nonconstant polynomial $t_1 p_1 \cdots p_N + 1$ would have an irreducible factor absent from the list. Only finitely many such irreducibles can divide the nonzero polynomial q, considered in $k[t_1, \dots, t_r]$. For example, any divisor from $k[t_1]$ must divide each coefficient of q when it is expanded in $t_2, \dots, t_r$, and a nonzero coefficient has finitely many irreducible divisors. Choose an irreducible $p(t_1)$ not dividing q.

The ideal (p) remains prime in $k[t_1,\dots,t_r]$, since its quotient is the polynomial ring in $t_2,\dots,t_r$ over the field $k[t_1]/(p)$. If p were invertible in A, we could write $ph/q^N = 1$ with h polynomial; the domain property gives $ph = q^N$. Primality would imply $p \mid q$, a contradiction. Thus A is not a field. This contradicts the preceding paragraph and proves $r = 0$.

All the original generators are therefore algebraic over k. If their respective algebraic degrees in the successive field extensions are $d_1,\dots,d_m$, the tower formula bounds $[L:k]$ by their finite product. Equivalently, repeatedly reducing powers gives a finite spanning set over k. Thus L/k is finite. □

The hypothesis is finite generation as an *algebra*, not merely as a field extension. The field $k(t)$ is generated by t as a field, but is not finitely generated as a k-algebra: no finite list of rational functions supplies inverses for every irreducible polynomial. That missing distinction is exactly what drives the proof.

17.3 Weak and strong forms of the Nullstellensatz

Theorem 17.2 (Weak Nullstellensatz)**.** *Every maximal ideal of $k[x_1,\dots,x_n]$ has the form*

$$\mathfrak{m}_a = (x_1 - a_1,\dots,x_n - a_n) \quad \textit{for a unique } a = (a_1,\dots,a_n) \in k^n.$$

Equivalently, every proper ideal has a common zero in k^n.

Proof. For a maximal ideal $\mathfrak{m}$, Zariski's lemma gives

$$L = P/\mathfrak{m} = k[\bar{x}_1,\dots,\bar{x}_n], \qquad [L:k] < \infty.$$

Every element of L has a linear minimal polynomial because k is algebraically closed. Hence $L = k$ through the given inclusion. Write $\bar{x}_i = a_i \in k$. Then

$$\mathfrak{m}_a = (x_1 - a_1,\dots,x_n - a_n) \subset \mathfrak{m}, \qquad P/\mathfrak{m}_a \cong k.$$

Thus $\mathfrak{m}_a$ is maximal and $\mathfrak{m} = \mathfrak{m}_a$. Uniqueness follows from $a_i = \bar{x}_i \in P/\mathfrak{m}$.

Every proper ideal lies in a maximal ideal, and its associated point annihilates the smaller ideal. Conversely, an ideal having a common zero cannot contain 1. □

Theorem 17.3 (Strong Nullstellensatz)**.** *For every ideal $J \subset P$, one has $I(V(J)) = \sqrt{J}$.*

Proof. We already proved $\sqrt{J} \subset I(V(J))$. Let $f \in I(V(J))$. If $f = 0$, the conclusion is immediate. Suppose $f \neq 0$ and introduce a new variable u. Consider

$$J' = (J, 1 - uf) \subset P[u].$$

For a common zero (a,b) of J', we would have

$$a \in V(J) \implies f(a) = 0, \qquad 1 - bf(a) = 0,$$

a contradiction. Hence $V(J') = \varnothing$, and the weak theorem gives $J' = P[u]$. Therefore

$$1 = \sum_{j=1}^{r} A_j(x,u) g_j(x) + B(x,u)(1 - uf(x)), \qquad g_j \in J.$$

Only finitely many $g_j \in J$ occur. In P_f, substitute $u = f^{-1}$:

$$1 = \sum_{j=1}^{r} A_j(x, f^{-1}) g_j.$$

Choose $N \geq 1$ clearing the powers of f in every denominator. Then

$$f^N = \sum_{j=1}^{r} h_j(x) g_j(x), \qquad h_j \in P.$$

This is an equality in P: the localization map is injective because P is a domain and $f \neq 0$. Thus $f^N \in J$, proving $f \in \sqrt{J}$. Introducing $1 - uf$ is called the *Rabinowitsch trick*; it turns the assertion that f is nonzero into a polynomial equation using one additional variable. □

Using the Nullstellensatz without forgetting its assumptions

Over an algebraically closed field, take $I = (x^2, xy) \subseteq k[x,y]$. The equations force $x = 0$, so

$$V(I) = V(x), \qquad I(V(I)) = (x), \qquad \sqrt{I} = (x).$$

The radical equality also follows purely algebraically: $x^2 \in I$ gives $(x) \subseteq \sqrt{I}$, while the prime ideal (x) contains I and hence contains its radical. But $I \neq (x)$, since $x \notin (x^2, xy)$. Therefore $k[x,y]/I$ is not the reduced coordinate ring of the set $V(I)$. It retains the nonzero nilpotent class of x, killed also by y.

More generally, the Nullstellensatz establishes a correspondence between radical ideals and closed sets, with inclusions reversed:

$$I \subseteq J \implies V(J) \subseteq V(I), \qquad Y \subseteq Z \implies I(Z) \subseteq I(Y).$$

It does not make all ideals recoverable from ordinary points. The ideals (x) and (x^2) already have identical zero sets.

The ground field cannot be dropped from the theorem. In $\mathbb{R}[x]$, the ideal (x^2+1) is maximal and radical, but its set of real zeros is empty. Hence its real vanishing ideal is all of $\mathbb{R}[x]$, not (x^2+1). Passing to $\mathbb{C}$ produces two zeros, i and $-i$, and restores the algebraically closed version. Alternatively $\operatorname{Spec} \mathbb{R}[x]$ retains the prime (x^2+1) as a point whose residue field is $\mathbb{C}$.

Why a function vanishing everywhere is nilpotent

The strong Nullstellensatz may be used as a nilpotence test. Let $A = k[x_1, \ldots, x_n]/I$ be a finite type algebra over algebraically closed k, without assuming it reduced. If $a \in A$ has zero image at every maximal ideal, lift it to a polynomial f. Then f vanishes on $V(I)$, so $f^N \in I$ for some N, and $a^N = 0$. Conversely every nilpotent vanishes in every residue field because a field has no nonzero nilpotents. Thus

$$\bigcap_{\mathfrak{m} \in \operatorname{MaxSpec} A} \mathfrak{m} = \sqrt{(0)}.$$

For a reduced finite type algebra, values at all closed points determine an element. For a nonreduced algebra they determine it only modulo the nilradical. The finite type and algebraically closed hypotheses are what make the preceding appeal to the Nullstellensatz possible.

17.4 Coordinate rings and irreducibility

The two operations now give an inclusion-reversing bijection

$$\{\text{algebraic subsets of } \mathbb{A}^n\} \longleftrightarrow \{\text{radical ideals of } P\}, \qquad X \longmapsto I(X), \quad J \longmapsto V(J).$$

The identity $V(I(X)) = X$ follows directly when X is algebraic; the other identity for radical ideals is the theorem.

Definition 17.4. The *coordinate ring* of an affine algebraic set $X \subset \mathbb{A}^n$ is $k[X] = P/I(X)$. Its elements are polynomial functions on X: two polynomials represent the same function exactly when their difference vanishes everywhere on X.

The ring $k[X]$ is reduced, and

$$X \text{ irreducible} \iff I(X) \text{ prime} \iff k[X] \text{ a domain.}$$

For the first implication, $fg \in I(X)$ gives

$$X = (X \cap V(f)) \cup (X \cap V(g)),$$

Irreducibility gives $X \subset V(f)$ or $X \subset V(g)$, hence $f \in I(X)$ or $g \in I(X)$. Conversely, for $X = Y \cup Z$ with Y, Z proper closed subsets, choose

$$f \in I(Y) \setminus I(X), \qquad g \in I(Z) \setminus I(X).$$

Such choices exist because $V(I(Y)) = Y$ and $V(I(Z)) = Z$. Their classes in $k[X]$ satisfy $\bar{f} \neq 0$, $\bar{g} \neq 0$, and $\bar{f}\bar{g} = 0$, so $k[X]$ is not a domain.

Maximal ideals of $k[X]$ are in bijection with points of X, by taking inverse images in P and applying the weak theorem. Write $\operatorname{MaxSpec} k[X]$ for the subset of maximal ideals in

Spec $k[X]$, with its subspace topology. The correspondence identifies X with MaxSpec $k[X]$ as a topological space: the equation $f(a) = 0$ is equivalent to $f \in \mathfrak{m}_a$, so closed sets correspond. It does *not* identify with all of Spec $k[X]$: the spectrum generally has additional nonclosed points, as in Lecture 16.

17.5 Exercises

Exercise 17.1. Let k be algebraically closed. Compute the radical and the zero set of $J = (x^2, xy) \subset k[x, y]$. Compute the coordinate ring of that zero set.

Exercise 17.2. Let k be algebraically closed. Find the irreducible components and coordinate ring of $V(xy) \subset \mathbb{A}^2_k$. Show explicitly that its coordinate ring is reduced but not a domain.

Exercise 17.3. Explain why algebraic closedness cannot be removed from the weak or strong Nullstellensatz as statements about k-valued points. Use $\mathbb{R}[x]$.

Exercise 17.4 (The equations of a union)**.** Let k be algebraically closed and $X = V(xy) \subseteq \mathbb{A}^2_k$. Determine its irreducible components and their intersection. Compute $I(X)$ and explain why the ideal (x^2y) defines the same point set but a different quotient ring.

Lecture 18
Regular maps, rational maps, and families

The correspondence between ideals and points extends to a correspondence between ring homomorphisms and maps of algebraic sets. Equations depending on parameters then provide examples of families. We also preview projective coordinates and the blowup of the plane. Reading: [2], Chapter I, Sections 3–4, for regular and rational maps, with later projective and blowup lectures supplying further details.

18.1 Regular functions and local rings

Let $X \subset \mathbb{A}^n$ be an affine variety over the algebraically closed field k, and let $U \subset X$ be open. A function $f : U \to k$ is *regular* if every point has an open neighborhood on which

$$f = g/h, \qquad g, h \in k[X], \qquad h \neq 0 \text{ at every point of that neighborhood.}$$

Thus division is allowed locally, where the denominator does not vanish. The set of regular functions on U is a ring, written $\mathcal{O}_X(U)$, under pointwise operations. To add fractions near a point, intersect their neighborhoods and use a common denominator. Products work the same way.

For example, $1/x$ is regular on $D(x) \subset \mathbb{A}^1$ but not at zero, whereas

$$\left.\frac{x^2}{x}\right|_{D(x)} = x|_{D(x)}$$

extends to all of $\mathbb{A}^1$. Thus the denominator in one presentation need not determine the maximal domain of regularity.

The *function field* of an affine variety is $k(X) = \mathrm{Frac}(k[X])$. This requires irreducibility, because we need a domain. Any fraction g/h with $h \neq 0$ defines a regular function on the nonempty open set $D_X(h)$. Two fractions defining the same function on a nonempty open set are equal in the fraction field. To see this, their difference has numerator q vanishing on a dense subset of X, so its closed zero set is all of X, and $q = 0$ in $k[X]$.

Conversely, a regular function on a nonempty open subset determines a unique element of $k(X)$. Choose a local fraction on one nonempty neighborhood. Every other local fraction agrees with it on their nonempty intersection, by irreducibility, so all give the same fraction-field element. This identifies a rational function with a regular function defined on some unspecified nonempty open subset, where shrinking the domain does not change it.

Definition 18.1. For $a \in X$, a *germ of a regular function at a* is an equivalence class of pairs (U, f) with U an open neighborhood of a and $f \in \mathcal{O}_X(U)$. The equivalence relation is

$$(U,f) \sim (V,g) \iff \exists \text{ open } W : a \in W \subset U \cap V,\ f|_W = g|_W.$$

Adding or multiplying representatives on an intersection defines a ring $\mathcal{O}_{X,a}$, the local ring at a. Write

$$\mathfrak{m}_a = \{g \in k[X] : g(a) = 0\}.$$

Evaluation at a gives $k[X]/\mathfrak{m}_a \cong k$, so $\mathfrak{m}_a$ is maximal.

Proposition 18.2. *For an affine variety X with coordinate ring $A = k[X]$,*

$$\mathcal{O}_X(X) = A, \qquad \mathcal{O}_{X,a} = A_{\mathfrak{m}_a}, \qquad k(X) = \operatorname{Frac}(A).$$

The local-ring identification sends the germ of g/h to that fraction in the localization $A_{\mathfrak{m}_a}$.

Proof. A fraction regular at a can be written locally with denominator not in $\mathfrak{m}_a$, so it lies in $A_{\mathfrak{m}_a}$. Every element of that localization has such a representative. Equality of germs agrees with equality in the fraction field by density, proving the local assertion.

For the global assertion, regard a global regular function f as an element of $\operatorname{Frac}(A)$ as above. It belongs to $A_{\mathfrak{m}}$ for every maximal ideal, because maximal ideals correspond to points. We prove

$$A = \bigcap_{\mathfrak{m} \text{ maximal}} A_{\mathfrak{m}} \subset \operatorname{Frac}(A).$$

For $z \in \operatorname{Frac}(A)$, set $J = \{a \in A : az \in A\}$, an ideal. Suppose $z \notin A$. Then $1 \notin J$, so $J \subset \mathfrak{m}$ for some maximal ideal $\mathfrak{m}$. Now

$$\begin{aligned} z \in A_{\mathfrak{m}} &\implies z = b/s \quad (b \in A,\ s \notin \mathfrak{m}) \\ &\implies sz = b \in A \\ &\implies s \in J \subset \mathfrak{m}, \end{aligned}$$

a contradiction. Hence $\bigcap_{\mathfrak{m}} A_{\mathfrak{m}} \subset A$; the reverse inclusion is immediate. Since polynomial functions are regular, $\mathcal{O}_X(X) = A$. The fraction-field assertion is the definition together with the preceding local-fraction identification. □

For affine algebraic sets $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$, their algebraic product is the set $X \times Y \subset \mathbb{A}^{n+m}$ with the equations and Zariski topology induced by these coordinates. This specifies the geometric product used below; its topology need not be the product topology of the two point sets.

For $0 \neq h \in A$, the maps

$$D_X(h) \longrightarrow \{(a,u) \in X \times \mathbb{A}^1 : h(a)u = 1\},$$
$$a \longmapsto (a, h(a)^{-1}), \qquad (a,u) \longmapsto a$$

are inverse functions, and their coordinates are regular on the indicated domains. The affine set on the right has coordinate ring $A[u]/(hu-1) \cong A_h$. The global assertion therefore gives

$$\mathscr{O}_X(D_X(h)) = A_h.$$

18.2 Morphisms and pullbacks

For affine algebraic sets $X \subset \mathbb{A}^n$, $Y \subset \mathbb{A}^m$, a *polynomial map* is a map of the form

$$\varphi(a) = \big(f_1(a), \ldots, f_m(a)\big), \qquad f_j \in k[X],$$

whose values lie in Y. We call these the morphisms of affine algebraic sets. For affine varieties, this is equivalent to requiring the coordinate functions to be regular, by the proposition. For open subsets of affine varieties, a morphism is a continuous map whose target coordinate functions pull back to regular functions. Equivalently, it is locally given by tuples of regular fractions; the continuity of such formulas follows in the next paragraph. An isomorphism is a morphism with an inverse that is also a morphism.

Write $V_X(g_\lambda) = X \cap V(g_\lambda)$ for a zero set inside X, and use the same convention for Y. For a polynomial map $\varphi = (f_1, \ldots, f_m)$,

$$\varphi^{-1}\big(V_Y(g_\lambda : \lambda \in \Lambda)\big) = V_X\big(g_\lambda(f_1, \ldots, f_m) : \lambda \in \Lambda\big).$$

Thus φ is Zariski continuous. For local fractions, clear denominators on each neighborhood to obtain the same conclusion there. The local statements imply continuity globally: if (U_i) is an open cover and $F \cap U_i$ is closed in U_i, then

$$X \setminus F = \bigcup_i \big(U_i \setminus (F \cap U_i)\big)$$

is open in X.

A morphism induces the *pullback* homomorphism

$$\varphi^* : k[Y] \longrightarrow k[X], \qquad g \longmapsto g \circ \varphi.$$

The arrow reverses because a function on the target can be evaluated after mapping into the target. If $X \xrightarrow{\varphi} Y \xrightarrow{\psi} Z$, then $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$.

Proposition 18.3. *For affine algebraic sets, there is a bijection*

$$\mathrm{Hom}(X,Y) \cong \mathrm{Hom}_{k\text{-Alg}}(k[Y], k[X]).$$

In particular, isomorphisms of affine algebraic sets correspond to k-algebra isomorphisms of their coordinate rings in the reverse direction.

Proof. Write $k[Y] = k[y_1,\dots,y_m]/I(Y)$. Given a homomorphism $\theta : k[Y] \to k[X]$, set $f_j = \theta(\bar{y}_j)$. Define $\varphi(a) = (f_1(a),\dots,f_m(a))$. If $g \in I(Y)$, then

$$g(f_1,\dots,f_m) = \theta(g(\bar{y}_1,\dots,\bar{y}_m)) = 0$$

in $k[X]$. Evaluating at a shows $\varphi(a) \in V(I(Y)) = Y$. Pulling back the coordinate functions recovers θ, and a map into affine space is determined by its coordinates, so the two constructions are inverse. Compatibility with composition proves the assertion about isomorphisms. □

For the parabola $Y = V(y - x^2)$, the inverse maps and corresponding ring isomorphism are

$$\mathbb{A}^1 \longrightarrow Y, \quad t \longmapsto (t,t^2), \qquad Y \longrightarrow \mathbb{A}^1, \quad (x,y) \longmapsto x,$$
$$k[x,y]/(y-x^2) \xrightarrow{\sim} k[t], \qquad \bar{x} \longmapsto t, \quad \bar{y} \longmapsto t^2.$$

By contrast, the cusp $C = V(y^2 - x^3)$ has a bijective morphism

$$\nu : \mathbb{A}^1 \longrightarrow C, \qquad t \longmapsto (t^2,t^3),$$

which is not an isomorphism. A nonzero cusp point has $x \neq 0$ and uniquely recovers $t = y/x$; the origin recovers $t = 0$. But the induced inclusion

$$k[t^2,t^3] \hookrightarrow k[t]$$

is not surjective, since every polynomial in t^2, t^3 has zero coefficient of t. The cusp ring equals $k[t^2,t^3]$: divide a polynomial by the monic polynomial $y^2 - x^3$ to get a remainder $a(x) + yb(x)$, and substituting (t^2,t^3) separates the even powers from the odd powers at least three, forcing both remainders to vanish. Thus even a bijection of geometric points need not preserve the full algebraic structure.

18.3 Rational maps and birational equivalence

A *rational map* $X \dashrightarrow Y$ between varieties is a morphism from a nonempty open subset of X to Y, where two representatives are identified if they agree on a nonempty open subset of their common domain. Irreducibility makes this an equivalence relation: the intersection of the open sets needed for transitivity remains nonempty. A rational map is *dominant* if its image is dense in Y; a morphism has the same terminology.

For affine varieties, write $I_Y(S) = \{g \in k[Y] : g|_S = 0\}$. Pullback satisfies

$$\ker \varphi^* = I_Y\big(\overline{\varphi(X)}\big).$$

Indeed, $g \circ \varphi = 0$ means g vanishes on $\varphi(X)$, equivalently on its closure. A proper closed subset of Y has a nonzero vanishing ideal in $k[Y]$. Therefore

$$\varphi \text{ dominant} \iff \ker \varphi^* = 0.$$

The injection extends to function fields by

$$k(Y) \hookrightarrow k(X), \qquad \frac{g}{h} \longmapsto \frac{\varphi^*(g)}{\varphi^*(h)}.$$

More generally, dominant rational maps of affine varieties correspond to k-embeddings $k(Y) \hookrightarrow k(X)$. To reconstruct a rational map, take finitely many coordinate generators of $k[Y]$, write their images as fractions on X, and restrict to the nonempty principal open where their common denominator is nonzero. The relations remain zero, so the tuple lands in Y. Injectivity makes its image dense by the same vanishing argument. The two constructions recover each other. Dominance is necessary: a constant map to $0 \in \mathbb{A}^1$ cannot pull back the rational function $1/t$.

Dominant rational maps can be composed: the inverse image of the nonempty open domain of the second map is nonempty because the first has dense image. Restricting to this inverse image gives the composite. Two varieties are *birational* if they have dominant rational maps in both directions whose composites agree with the identity on nonempty open subsets. The construction just given shows that affine varieties are birational exactly when their function fields are isomorphic over k. Their coordinate rings may still differ. The cusp and the line are birational, with inverse $t = y/x$ on $x \neq 0$, although they are not isomorphic. Rational maps therefore describe the part of geometry visible on dense open sets, and may forget special points.

Dominance and injectivity of pullback

Let $f : X \to Y$ be a morphism of affine varieties over an algebraically closed field. It is called dominant if its image is dense in Y. Then

$$f \text{ dominant} \quad \Longleftrightarrow \quad f^* : k[Y] \longrightarrow k[X] \text{ injective.}$$

If $g \in \ker f^*$, then g vanishes on $f(X)$. For a dominant map it vanishes on its closure Y, so $g = 0$ in $k[Y]$. Conversely, if the image is not dense, its closure is a proper closed subset of Y. Some nonzero element of $k[Y]$ vanishes on that subset, so its pullback is zero. This contradicts injectivity.

For a dominant map, the pullback extends to an embedding of function fields by $g/h \mapsto f^*g/f^*h$. The denominator is nonzero because f^* is injective. Without dominance this need not be possible: the inclusion of the origin into $\mathbb{A}^1$ sends x to zero, so it cannot assign an image to $1/x$ in a field.

Birational does not mean isomorphic

For the cusp parametrization $\nu(t) = (t^2, t^3)$, the pullback is

$$k[x,y]/(y^2-x^3) \hookrightarrow k[t], \qquad x \mapsto t^2, \quad y \mapsto t^3.$$

It is injective but not surjective, since its image has no nonzero linear term in t. Its extension to fraction fields is an isomorphism because $t = y/x$. Thus ν is birational. On $D(x)$ the inverse is the regular function y/x, while at the cusp this fraction is not regular. The map is also bijective on ordinary points, so bijectivity and birationality together still do not guarantee an isomorphism.

Contrast the map $\mathbb{A}^1 \to \mathbb{A}^1$, $t \mapsto t^2$ over a field of characteristic different from two. It is dominant, but its function-field inclusion $k(x) \hookrightarrow k(t)$ has degree two; a generic value has two preimages after extending scalars. It is not birational. Over characteristic $p > 0$, the map $t \mapsto t^p$ is bijective on points of an algebraically closed field, but its pullback misses t and its function-field degree is p. This is another reason to test maps on rings rather than count preimages alone.

18.4 Families and fiber rings

A *fiber* of a map $f : X \to S$ at a point s is the inverse image $f^{-1}(\{s\})$. An algebraic family is a morphism $X \to S$, with S regarded as a space of parameters. A ring map $B \to A$ gives the continuous map $\operatorname{Spec} A \to \operatorname{Spec} B$ already constructed in Lecture 16. At a point s corresponding to $\mathfrak{p} \subset B$, the *fiber ring* is

$$A \otimes_B \kappa(s).$$

This ring retains the equations of the fiber, including nilpotents. Put $S = B \setminus \mathfrak{p}$. By tensoring with a quotient and a localization,

$$A \otimes_B \kappa(s) \cong S^{-1}(A/\mathfrak{p}A).$$

A prime of the right side corresponds to a prime $\mathfrak{q}$ of A satisfying

$$\mathfrak{p}A \subset \mathfrak{q}, \qquad \mathfrak{q} \cap S = \varnothing, \qquad \text{equivalently} \quad \mathfrak{q} \cap B = \mathfrak{p}.$$

Here intersections with B mean inverse images under the given map $B \to A$. These are exactly the prime correspondences for quotients and localization. For a closed k-point with maximal ideal $\mathfrak{m}$ and residue field k,

$$A \otimes_B k \cong A/\mathfrak{m}A.$$

Take $k[t] \to k[x,y]$, $t \mapsto xy$. The associated map of classical points is $(x,y) \mapsto xy$. The fiber over $a \in k$ has ring

$$k[x,y]/(xy-a).$$

For $a \neq 0$, x is invertible and $y = a/x$, so this is $k[x,x^{-1}]$: the fiber is a punctured line. For $a = 0$, the fiber is the union of two crossing axes. Over the generic point (0) of $\operatorname{Spec} k[t]$, its residue field is $k(t)$ and the fiber ring is

$$k(t)[x,y]/(xy-t) \cong k(t)[x,x^{-1}].$$

The generic fiber is a variety over the field $k(t)$, not the fiber over a mysterious scalar value in k.

Fibers can acquire nilpotents even when the total space is reduced. For the map $\mathbb{A}^1 \to \mathbb{A}^1$, $x \mapsto x^2$, the fiber ring at zero is $k[x]/(x^2)$. Its classical zero set is a single point, whose reduced coordinate ring is k; the fiber ring remembers the square in the equation. In characteristic different from two, a nonzero fiber instead has two distinct points because $x^2 - a$ has two distinct roots. We retain the entire quotient ring when recording a fiber algebraically. To recover only its classical algebraic set, we replace the defining ideal by its radical. The terminology of scheme-theoretic fibers will be introduced after schemes have been defined in Lecture 25.

A fiber is formed over the residue field of the base point

For $A \to B$ and $\mathfrak{p} \in \operatorname{Spec} A$, let $S = A \setminus \mathfrak{p}$. The fiber ring has two equivalent descriptions:

$$B \otimes_A \kappa(\mathfrak{p}) \cong S^{-1}B/\mathfrak{p}S^{-1}B, \qquad \kappa(\mathfrak{p}) = \operatorname{Frac}(A/\mathfrak{p}).$$

Here $\kappa(\mathfrak{p})$ is identified with $A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$. First invert elements outside $\mathfrak{p}$, then impose that elements of $\mathfrak{p}$ vanish. Its prime ideals correspond exactly to primes $\mathfrak{q} \subseteq B$ with $\mathfrak{q} \cap A = \mathfrak{p}$: avoiding S gives $\mathfrak{q} \cap A \subseteq \mathfrak{p}$, and containing $\mathfrak{p}B$ gives the reverse inclusion.

For the family $xy = t$, take $B = k[t,x,y]/(xy-t)$. At $t = a \neq 0$, the fiber is

$$k[x,y]/(xy-a) \cong k[x,x^{-1}], \qquad y \longmapsto a/x.$$

At $a = 0$ it is $k[x,y]/(xy)$, the union of two axes. The total space is nevertheless isomorphic to $\mathbb{A}^2$, since eliminating t gives $B \cong k[x,y]$. A nonsingular total space can therefore have a singular fiber. The singularity belongs to the fiber equation and to the map to the base, not merely to the total space considered on its own.

18.5 Projective directions and an incidence variety

Projective space $\mathbb{P}^n(k)$ is the set of one-dimensional vector subspaces of k^{n+1}. A nonzero vector $(a_0,\dots,a_n)$ represents the point $[a_0 : \cdots : a_n]$, and multiplying all coordinates by the same nonzero scalar does not change the point. On the chart $a_i \neq 0$, divide all coordinates by a_i to obtain n affine coordinates. Lecture 19 develops its equations and topology.

The projective line parametrizes directions through the origin in the affine plane. Away from the origin there is a direction map

$$\mathbb{A}^2 \setminus \{(0,0)\} \longrightarrow \mathbb{P}^1, \qquad (x,y) \longmapsto [x : y].$$

To keep a direction even over the origin, consider the incidence equation

$$\widetilde{\mathbb{A}^2} = \{((x,y),[u:v]) \in \mathbb{A}^2 \times \mathbb{P}^1 : xv = yu\}.$$

This is the *blowup of the plane at the origin*. Projection to (x,y) has a unique point over every nonzero point, but its fiber over the origin is the whole $\mathbb{P}^1$, called the *exceptional fiber*. The origin has been replaced by its possible directions. Away from the origin the incidence set is the *graph* of the direction map: for a map $f : X \to Y$, its graph is the subset $\{(x, f(x)) : x \in X\} \subset X \times Y$.

The two charts and their projections are

chart	coordinate ratio	equation	projection
$u \neq 0$	$s = v/u$	$y = xs$	$(x,s) \longmapsto (x,xs)$
$v \neq 0$	$r = u/v$	$x = yr$	$(r,y) \longmapsto (ry,y)$

Each chart is an affine plane. On the overlap,

$$r = 1/s, \qquad y = xs.$$

These formulas are regular where $s \neq 0$ and explain how two affine planes fit together. They also show that the inverse away from the origin is regular on $x \neq 0$ or $y \neq 0$. Lecture 22 studies this construction, transformed curves, and its relation to ideals in detail.

18.6 Exercises

Exercise 18.1. Show that the hyperbola $H = V(xy-1)$ is isomorphic to $D(t) \subset \mathbb{A}^1$. Compute its coordinate ring and function field.

Exercise 18.2. For the family $xy = t$, compute the fiber of the closed subset $x = y$ over $a \in k$. Explain the difference at $a = 0$ between the fiber ring and its classical point set.

Exercise 18.3. In the blowup chart (x,s) with $y = xs$, find the limit direction of the punctured line $y = mx$ as it approaches the origin. Compare it with the punctured parabola $y = x^2$.

Lecture 19
Projective varieties and the meaning of intersection

Projective geometry can be calculated in ordinary affine coordinates. Intersections, however, require quotient rings that retain more information than their underlying points. We will construct a projective conic explicitly and introduce intersection multiplicity through small quotient rings. Reading: [2], Chapter I, §§2–3 and the intersection-multiplicity exercise in §5.

19.1 Projective coordinates

Two distinct lines in the affine plane need not meet: the lines $y = 0$ and $y = 1$ are parallel. A polynomial equation can also appear to lose solutions when its leading coefficient changes. Projective coordinates supply additional points at which this missing behaviour can be recorded algebraically. They do not make every geometric problem trivial; they give a larger space in which degree and intersection have cleaner rules.

Definition 19.1. The projective space $\mathbb{P}^n_k$ is the set of equivalence classes of nonzero vectors in k^{n+1}, where

$$(a_0, \dots, a_n) \sim (\lambda a_0, \dots, \lambda a_n) \quad \text{for } \lambda \in k^\times.$$

The class of a vector is written $[a_0 : \cdots : a_n]$. These are its *homogeneous coordinates*.

Equivalently,

$$\mathbb{P}^n_k = \{\ell \subset k^{n+1} : \ell \text{ is a one-dimensional subspace}\}, \qquad [a_0 : \cdots : a_n] \longleftrightarrow k(a_0, \dots, a_n).$$

The zero vector spans no line. Rescaling is simultaneous: over $\mathbb{Q}$,

$$[1 : 2 : 3] = [2 : 4 : 6], \qquad [1 : 2 : 3] \neq [1 : 4 : 3].$$

Every point of $\mathbb{P}^1$ with first coordinate nonzero has the unique form $[1 : t]$. There is one further point, $[0 : 1]$. Accordingly,

$$\mathbb{P}^1 = \{[1 : t] : t \in k\} \cup \{[0 : 1]\}.$$

The extra point is called the point at infinity for this choice of affine chart. It is not intrinsically different from the other points: exchanging the coordinates moves it into an affine chart.

19.2 Equations and affine charts

Recall that a monomial $X_0^{a_0}\cdots X_n^{a_n}$ has total degree $a_0+\cdots+a_n$. A polynomial is *homogeneous of degree d* if all its nonzero monomials have degree d. An ideal is *homogeneous* if it is generated by homogeneous polynomials; equivalently, it contains every homogeneous part of each of its elements. For example, $(XZ-Y^2)$ is homogeneous, whereas $(Y-X^2)$ is not.

A general polynomial in $X_0,\dots,X_n$ cannot be evaluated unambiguously on a projective point. Even the value of a homogeneous polynomial changes when coordinates are rescaled. What is well defined is its vanishing: if F is homogeneous of degree d, then

$$F(\lambda a_0,\dots,\lambda a_n)=\lambda^d F(a_0,\dots,a_n).$$

Consequently $F(a)=0$ is independent of the representative.

For a homogeneous ideal $I\subset S=k[X_0,\dots,X_n]$, define

$$V_+(I)=\{[a_0:\cdots:a_n]:F(a_0,\dots,a_n)=0 \text{ for every homogeneous } F\in I\}.$$

These sets are called *projective algebraic sets*. They satisfy the same finite-union and arbitrary-intersection identities as affine zero sets, so they are the closed sets of a topology, the *projective Zariski topology*. A *projective variety* is a nonempty irreducible projective algebraic set, with polynomial relations taken to be all relations vanishing on it. Its affine coordinate rings are therefore reduced. A *quasiprojective variety* is an open subset of a projective variety. A subset is *locally closed* if it is the intersection of an open and a closed subset; equivalently, it is open in its closure. Thus an irreducible locally closed subset of projective space is quasiprojective.

Set $U_i=\{X_i\neq 0\}$. The maps

$$U_i\longrightarrow \mathbb{A}^n,\qquad [X_0:\cdots:X_n]\longmapsto\left(\frac{X_0}{X_i},\dots,\widehat{\frac{X_i}{X_i}},\dots,\frac{X_n}{X_i}\right)$$

are coordinate identifications; the hat means that the indicated coordinate is omitted. Their inverse inserts 1 in the ith position. The sets U_i cover projective space because not all coordinates vanish.

On U_i, divide a homogeneous equation of degree d by X_i^d:

$$F(X_0,\dots,X_n)=0 \quad\Longleftrightarrow\quad F\left(\frac{X_0}{X_i},\dots,1,\dots,\frac{X_n}{X_i}\right)=0.$$

Conversely, multiplying an affine equation by a sufficiently large power of X_i makes it homogeneous. On $U_i \cap U_j$, the coordinate changes are

$$\frac{X_\ell}{X_j} = \frac{X_\ell/X_i}{X_j/X_i}, \qquad X_j/X_i \neq 0.$$

Thus the affine equations and their regular coordinate changes agree.

For example, on $\mathbb{P}^1$ use $t = X_1/X_0$ on U_0 and $u = X_0/X_1$ on U_1. Their overlap has $t \neq 0$, and

$$u = t^{-1}.$$

This is the simplest example of a space assembled from affine pieces.

For a projective algebraic set X, let $I_+(X)$ be the ideal generated by all homogeneous polynomials vanishing on X. The ring $R = S/I_+(X)$ is its *homogeneous coordinate ring*. Its grading is inherited from S. This ring differs from a ring of globally defined functions on X: a homogeneous polynomial of positive degree changes under rescaling. After inverting X_i, declare $\deg X_i^{-1} = -1$; then $\deg(F/X_i^d) = \deg F - d$. The chart $X_i \neq 0$ has coordinate ring

$$\bigl(R[X_i^{-1}]\bigr)_0.$$

Here the subscript means the degree-zero part: a typical element is F/X_i^d with F homogeneous of degree d. The degree-zero condition is precisely what makes the expression invariant under simultaneous rescaling. The entire ring $R[X_i^{-1}]$ would still retain an unnecessary scaling coordinate.

A function on an open subset of a projective variety is *regular* if its expression on every affine chart is locally a fraction of polynomials with denominator nonzero. A *morphism* between quasiprojective varieties is a continuous map such that, on the inverse image of each target affine chart, its coordinate functions are regular. This agrees with the affine definition in Lecture 18. Because chart transitions are regular in both directions, using a different set of charts gives the same condition. An isomorphism is a morphism with a morphism as inverse.

19.3 Homogenization and projective closure

The *projective closure* of $X \subset \mathbb{A}^n$ is its closure in $\mathbb{P}^n$ after the embedding $(a_1, \dots, a_n) \mapsto [1 : a_1 : \cdots : a_n]$. For $f \in k[x_1, \dots, x_n]$ of total degree d, its *homogenization* is

$$f^{\mathrm{h}}(X_0, \dots, X_n) = X_0^d f(X_1/X_0, \dots, X_n/X_0).$$

The right side is a polynomial because multiplication by X_0^d clears every denominator. For example, $y - x^2$ becomes $X_0 Y - X^2$.

If an affine variety has ideal I, its projective closure has homogeneous ideal generated by f^{h} for *all* $f \in I$. It is tempting to homogenize an arbitrary chosen generating set and stop.

This can introduce extra components at infinity, because taking polynomial combinations and homogenizing do not commute without additional powers of X_0.

Here is a concrete warning. The affine twisted cubic

$$(x,y,z) = (t,t^2,t^3)$$

has ideal $(y-x^2, z-xy)$. With projective coordinates $[W:X:Y:Z]$, homogenizing these two generators gives

$$WY - X^2 = 0, \qquad WZ - XY = 0.$$

On $W = 0$, both equations hold along the whole line $W = X = 0$. But

$$x(z-xy) - y(y-x^2) = xz - y^2$$

is also in the affine ideal. Its homogenization $XZ - Y^2$ does not vanish everywhere on that line; in particular it excludes $[0:0:1:0]$. The naive two-equation answer is too large.

Definition 19.2. For an ideal J and an element w, its saturation by w is

$$J : w^\infty = \{g : \text{there is } r \geq 0 \text{ with } w^r g \in J\}.$$

Equivalently, it is the contraction of J from the localization in which w is invertible.

If J is generated by homogenizations of generators of I, then $J : X_0^\infty$ is the ideal obtained by homogenizing all of I. To understand this statement, dehomogenize a polynomial relation in I and then homogenize it to one sufficiently large common degree. The resulting relation says that a power of X_0 times the desired homogeneous polynomial lies in J. Conversely, membership after multiplying by X_0^r becomes membership in I after setting $X_0 = 1$. This is also why localization and contraction give the correct closure.

Caution 19.3. For an ideal K, define $J : K^r = \{g : gK^r \subset J\}$ and $J : K^\infty = \bigcup_{r\geq 0}(J : K^r)$. The ideal $S_+ = (X_0, \ldots, X_n)$ is called the *irrelevant ideal*; its affine zero set is only the zero vector, which gives no projective point. Saturation by X_0 removes equations' components absent from the chosen chart $X_0 \neq 0$. Saturation by S_+ preserves the localized homogeneous equations on *every* projective chart. For example, $J = (X_0)$ satisfies $J : X_0^\infty = S$ but $J : S_+^\infty = J$ when $n \geq 1$: in the latter assertion, $gX_1^r \in (X_0)$ forces $g \in (X_0)$. Thus the two operations differ. Taking a radical is another operation; it removes nilpotents from the quotient ring.

The hyperbola and its two points at infinity

The affine equation $xy = 1$ homogenizes to $XY = Z^2$ in $\mathbb{P}^2$. On the chart $Z = 1$ we recover the original hyperbola. On $Z = 0$ the equation gives $XY = 0$, so the additional points are

$$P = [1:0:0], \qquad Q = [0:1:0].$$

Near P, use $X = 1$ and coordinates $u = Y/X$, $v = Z/X$. The equation becomes $u = v^2$, so v is a local parameter and the curve is nonsingular there. Near Q the symmetric calculation

gives the same conclusion. The line at infinity $Z = 0$ has equation $v = 0$ in each chart and meets the curve with multiplicity one at each of P, Q.

The parametrization

$$[s : t] \longmapsto [s^2 : t^2 : st]$$

identifies this projective conic with $\mathbb{P}^1$. The chart formulas used for the Veronese conic give its inverse. On $st \neq 0$, the affine functions are $x = s/t$, $y = t/s$, so the original hyperbola corresponds to $\mathbb{P}^1$ with the two points $[1 : 0]$ and $[0 : 1]$ removed. The calculation explains which points are missing from the affine model, rather than treating infinity as a single universal point.

How to calculate a projective chart ring

For $R = k[X, Y, Z]/(XY - Z^2)$, invert X and take degree zero. It is generated by $u = Y/X$ and $v = Z/X$, with relation $u - v^2 = 0$:

$$(R[X^{-1}])_0 \cong k[u, v]/(u - v^2) \cong k[v].$$

To check generation, every homogeneous monomial of degree d divided by X^d becomes a product of Y/X and Z/X. To check the relation, divide $XY - Z^2$ by X^2. Finally the map $v \mapsto Z/X$ has inverse obtained from $X = 1$, $Y = v^2$, $Z = v$, which proves that no further relations have been overlooked. Merely substituting $X = 1$ into a picture is justified by this degree-zero calculation.

19.4 Maps to projective space

Suppose $F_0, \ldots, F_m$ are homogeneous polynomials of the same degree on a projective variety X, and have no common zero on X. They define

$$X \longrightarrow \mathbb{P}^m, \qquad P \longmapsto [F_0(P) : \cdots : F_m(P)].$$

The common degree makes the result independent of coordinates. The absence of a common zero ensures an actual projective point at every input. On the open set $F_i \neq 0$, target affine coordinates are the regular functions F_j/F_i. This proves that the displayed map is a morphism.

The quadratic Veronese map is

$$\nu_2 : \mathbb{P}^1 \longrightarrow \mathbb{P}^2, \qquad [s : t] \longmapsto [s^2 : st : t^2].$$

The three entries cannot vanish together. Its image lies on $XZ - Y^2 = 0$. On the chart $X \neq 0$ of that conic, define the inverse by $[X : Y : Z] \mapsto [X : Y]$; on $Z \neq 0$, use $[Y : Z]$. On their overlap these agree because $XZ = Y^2$. Direct substitution shows that both compositions are

identities. Thus ν_2 identifies $\mathbb{P}^1$ with the conic, with its full algebraic structure, rather than merely giving a bijection of points.

A *closed embedding* is a morphism that is an isomorphism onto a closed subvariety of its target. The Veronese map is a closed embedding by the explicit inverse above. It represents the abstract curve $\mathbb{P}^1$ by a quadratic equation, whereas a line in $\mathbb{P}^2$ is given by a linear equation. Lecture 20 defines the degree of an embedded variety and quantifies this distinction.

19.5 Local intersection multiplicity

A *plane curve* in this section is the zero set of one nonconstant polynomial in $\mathbb{A}^2$, or one nonzero homogeneous polynomial of positive degree in $\mathbb{P}^2$. It may have several irreducible components; the equations may also specify repeated factors. Let two such curves have local equations $f = 0$ and $g = 0$ near a common point $p = (a,b)$. Write $\mathfrak{m}_p = (x-a, y-b)$ for the point's maximal ideal in $k[x,y]$. Their intersection is *isolated at p* if some open neighborhood of p meets the common zero set only at p. Their local intersection algebra is

$$B_p = k[x,y]_{\mathfrak{m}_p}/(f,g).$$

Isolation means that this local quotient has only one prime ideal, its maximal ideal. It is Noetherian as a quotient of a localized polynomial ring, hence Artinian by the dimension-zero criterion in Lecture 9. Define the *intersection multiplicity* by

$$i_p(f,g) = \operatorname{length}(B_p) = \dim_k B_p.$$

The equality holds here because k is algebraically closed and every simple module over this local ring is its residue field k. A composition series therefore counts precisely its vector-space dimension. The same definition works in any affine chart of $\mathbb{P}^2$; changing charts changes local equations by invertible factors and gives isomorphic local quotient rings.

For the line $y = 0$ and parabola $y = x^2$ at the origin,

$$B_0 = k[x,y]_{(x,y)}/(y, y-x^2) \cong k[x]_{(x)}/(x^2).$$

Hence

$$B_0 = k \cdot 1 \oplus k \cdot \bar{x}, \qquad \bar{x} \neq 0, \quad \bar{x}^2 = 0, \qquad i_0(y, y-x^2) = 2.$$

The nilpotent $\bar{x}$ records the second order of contact at the single intersection point.

For later computations, the *formal partial derivatives* of a polynomial are defined on monomials by

$$\frac{\partial (x^i y^j)}{\partial x} = i x^{i-1} y^j, \qquad \frac{\partial (x^i y^j)}{\partial y} = j x^i y^{j-1},$$

and extended by addition and multiplication by scalars. The integer coefficients are interpreted in k, so they may vanish in positive characteristic. Write f_x, f_y for the resulting derivatives.

To define the tangent terminology used here, translate p to the origin and write $f = f_1 + f_2 + \cdots$, where f_j is homogeneous of degree j. The *linear part* is $f_1 = \alpha x + \beta y$. If $(\alpha, \beta) \neq (0,0)$, the line $\alpha u + \beta v = 0$ in the vector space of directions is the *tangent line* to $f = 0$ at p. Two curves with nonzero independent linear parts meet *transversely*: their tangent lines are distinct. Lecture 21 identifies this linear calculation with the intrinsic tangent space defined from the local ring.

For the coordinate axes, $B_0 = k[x,y]_{(x,y)}/(x,y) = k$, so $i_0(x,y) = 1$. More generally, independent linear parts give

$$\mathfrak{m}_p = (f,g) + \mathfrak{m}_p^2.$$

Apply Nakayama's lemma to $M = \mathfrak{m}_p/(f,g)$:

$$M = \mathfrak{m}_p M \implies M = 0 \implies B_p = k \implies i_p(f,g) = 1.$$

This is the algebraic meaning of a transverse intersection.

19.6 Exercises

Exercise 19.1. Describe the conic $C = V_+(XZ - Y^2)$ on the two charts $X \neq 0$ and $Z \neq 0$. Check that these charts cover C.

Exercise 19.2. Find the intersection in $\mathbb{P}^2$, with coordinates $[X : Y : Z]$, of the projective closures of $y = 0$ and $y = 1$ in the chart $Z \neq 0$.

Exercise 19.3. Homogenize $xy - 1$ and find all points at infinity of its projective closure. Explain why there are two.

Exercise 19.4. Does $[X : Y : Z] \mapsto [X : Y]$ define a morphism from all of $\mathbb{P}^2$ to $\mathbb{P}^1$? Can the missing value be supplied to make it a morphism?

Exercise 19.5. Compute the intersection multiplicity at the origin of $y = 0$ and $y = x^r$, for an integer $r \geq 1$. Explain the difference between $r = 1$ and $r > 1$.

Lecture 20
Hilbert polynomials and Bézout's theorem

The plane Bézout theorem will be derived from exact sequences of graded modules. We will calculate the Hilbert polynomial of a hypersurface, prove the plane form of Bézout's theorem with local multiplicities, and use examples to see exactly why its assumptions matter. Reading: [2], Chapter I, §7, together with Lecture 11.

20.1 Hilbert functions and degree

Let $S = k[X_0, \ldots, X_n]$ with every variable in degree 1. A monomial of degree m is determined by nonnegative exponents whose sum is m. To count these exponent tuples, arrange m identical symbols and n separators in a row; the numbers of symbols before, between, and after the separators are the $n+1$ exponents. Choosing the n separator positions among $m+n$ positions gives

$$\dim_k S_m = \binom{m+n}{n} \qquad (m \geq 0).$$

We set $S_m = 0$ for $m < 0$. For a finitely generated graded S-module M, recall its Hilbert function and Hilbert series:

$$h_M(m) = \dim_k M_m, \qquad H_M(t) = \sum_m h_M(m) t^m.$$

The Hilbert polynomial $P_M(m)$ agrees with $h_M(m)$ for all sufficiently large integers m. Its existence was established in Lecture 11; it is not generally a formula for every small degree.

Recall that the dimension of a topological space is the supremum of the lengths r of strict chains $Z_0 \subsetneq \cdots \subsetneq Z_r$ of nonempty irreducible closed subsets. For an affine algebraic set this is the Krull dimension of its coordinate ring, by the prime-ideal correspondence. For projective algebraic sets it is computed on affine charts. For a nonempty projective algebraic set X of dimension r, the Hilbert polynomial of its homogeneous coordinate ring has degree r. Its leading term has the form

$$P_X(m) = \frac{\deg X}{r!} m^r + \text{terms of lower degree.}$$

This defines the degree of the embedded algebraic set. The connection between polynomial growth and dimension comes from the graded dimension theorem reviewed in Lecture 12. We now compute these polynomials without assuming an intersection theorem.

A short exact sequence of graded modules is exact in each degree:

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0 \quad \Longrightarrow \quad 0 \longrightarrow M'_m \longrightarrow M_m \longrightarrow M''_m \longrightarrow 0.$$

Taking dimensions, generating series, and eventual polynomials gives

$$\begin{aligned} h_M(m) &= h_{M'}(m) + h_{M''}(m), \\ H_M(t) &= H_{M'}(t) + H_{M''}(t), \\ P_M(m) &= P_{M'}(m) + P_{M''}(m). \end{aligned}$$

20.2 The Hilbert polynomial of a hypersurface

A *projective hypersurface* is the zero set of one nonzero homogeneous polynomial of positive degree. Let $F \neq 0$ be homogeneous of degree $d > 0$. Multiplication by F is injective on the polynomial ring, since it is a domain. To make the map preserve degree, use the shifted graded module $S(-d)$, defined by $S(-d)_m = S_{m-d}$. We obtain

$$0 \longrightarrow S(-d) \xrightarrow{F} S \longrightarrow S/(F) \longrightarrow 0.$$

Consequently

$$H_{S/(F)}(t) = \frac{1-t^d}{(1-t)^{n+1}}, \qquad P_{S/(F)}(m) = \binom{m+n}{n} - \binom{m-d+n}{n}.$$

The binomial expressions in the polynomial formula are polynomials in m. The exact Hilbert function instead uses $S_{m-d} = 0$ when $m < d$. Keeping this distinction avoids incorrect negative-degree counts.

The two degree-n leading terms cancel. The new leading term is

$$\frac{d}{(n-1)!} m^{n-1}.$$

Thus a degree-d hypersurface has projective dimension $n-1$ and degree d. If F has repeated factors, this is the degree of the hypersurface with the multiplicities specified by its equation; replacing (F) by its radical would describe a different object.

20.3 The Hilbert series of a plane intersection

Work in $S = k[X,Y,Z]$. Let F, G be nonzero homogeneous polynomials of positive degrees d, e, with no common nonconstant factor. Equivalently, the plane curves they define have no common irreducible component. The polynomials themselves may have repeated factors.

The basic algebraic point is that multiplication by G is injective on $S/(F)$. Indeed, if $GH \in (F)$, unique factorization and the absence of a common factor imply $F \mid H$. This argument remains valid when F is reducible or has repeated factors: compare the exponent of each irreducible factor of F.

It follows that

$$0 \longrightarrow (S/(F))(-e) \xrightarrow{G} S/(F) \longrightarrow R \longrightarrow 0, \qquad R = S/(F,G),$$

is exact. Combining this with the one-equation calculation yields

$$\begin{aligned} H_R(t) &= \frac{(1-t^d)(1-t^e)}{(1-t)^3} \\ &= \frac{(1+t+\cdots+t^{d-1})(1+t+\cdots+t^{e-1})}{1-t}. \end{aligned}$$

If the numerator is $Q(t)$, its degree is $d+e-2$. For every $m \geq d+e-2$, the coefficient of t^m in $Q(t)/(1-t)$ is the sum of all coefficients of Q, namely

$$Q(1) = de.$$

We have proved directly that $h_R(m) = de$ for all such m. We still need to show why this number is the sum of the local intersection multiplicities. That bridge is essential; naming de the answer would not yet prove Bézout's geometric statement.

20.4 The Hilbert polynomial and total local length

First, $V_+(F,G)$ is a finite set. Here is how the dimension results apply. If the intersection contained an irreducible projective curve, that curve would have a height-one homogeneous prime ideal in $k[X,Y,Z]$. Height-one primes in this unique factorization domain are generated by irreducible polynomials. Its generator would divide both F and G, contrary to the hypothesis. There is no component of larger dimension either, since the nonzero equation F already cuts out dimension 1. A zero-dimensional projective algebraic set has only finitely many points, by Noetherianity and the dimension-zero case of the Nullstellensatz on affine charts.

Choose a linear form L nonzero at every intersection point. This is possible over our infinite field k: each point excludes one proper linear subspace of the space of linear forms, and a finite union of proper linear subspaces cannot fill that space. Change homogeneous coordinates so that $L = Z$. Thus all intersection points lie in $Z \neq 0$.

Let $f(x,y) = F(x,y,1)$ and $g(x,y) = G(x,y,1)$. The entire intersection is represented on this chart by the finite-dimensional algebra

$$B = k[x,y]/(f,g) \cong \left(R[Z^{-1}]\right)_0.$$

We justify its dimension directly rather than assuming a length formula. Since $V_+(F,G,Z)$ is empty, the affine zero set of the homogeneous ideal (F,G,Z) in $\mathbb{A}^3$ is only the origin. The affine Nullstellensatz says that its radical is (X,Y,Z). Thus sufficiently large powers of X,Y,Z lie in this ideal. Only finitely many monomials survive in $S/(F,G,Z)$, so

$$(R/ZR)_m = 0 \qquad \text{for sufficiently large } m.$$

Equivalently, multiplication $Z : R_m \to R_{m+1}$ is eventually surjective. We already calculated that both source and target eventually have dimension de, so these maps are eventually isomorphisms.

On the other hand, each element of $(R[Z^{-1}])_0$ is represented by a/Z^m with $a \in R_m$. Increasing m replaces a by Za. Therefore this ring, as a vector space, is the direct limit

$$R_0 \xrightarrow{Z} R_1 \xrightarrow{Z} R_2 \xrightarrow{Z} \cdots.$$

In this direct limit, a vector is represented at some stage R_m, and two representatives are equal if their images agree at a common later stage. Since the transition maps are isomorphisms from some stage onward, every representative can be compared in that stable part and the limit has dimension de. This also proves directly that B is finite-dimensional.

Finally, a finite-dimensional commutative k-algebra decomposes as a product of its local Artinian rings at its finitely many maximal ideals:

$$B \cong \prod_{p \in V(f,g)} B_{\mathfrak{m}_p}.$$

Recall the algebra behind this statement from Lecture 9: the nilradical of an Artinian ring is nilpotent, sufficiently large powers of distinct maximal ideals are comaximal, and the Chinese remainder theorem separates the resulting local factors. Each local factor has residue field k by the Nullstellensatz. Its vector-space dimension is therefore its module length. Taking dimensions proves

$$\dim_k B = \sum_p \operatorname{length}(B_{\mathfrak{m}_p}) = \sum_p i_p(F,G).$$

Theorem 20.1 (Plane Bézout theorem)**.** *Let $F,G \in k[X,Y,Z]$ be homogeneous of positive degrees d,e, over an algebraically closed field. If they have no common nonconstant factor, their projective intersection is finite and*

$$\boxed{\sum_{p \in V_+(F,G)} i_p(F,G) = de.}$$

Here the multiplicity is the length of the local quotient by their equations in any affine chart containing p.

Proof. The preceding graded exact sequence calculates the eventual Hilbert function as *de*. The affine-chart argument identifies this eventual value with the sum of local lengths. Each step used the no-common-component hypothesis or an explicitly stated property of the ground field, so it proves precisely the asserted theorem. □

20.5 Examples of Bézout's theorem

Take the conic $YZ-X^2=0$ and the line $Y=0$. Their only projective intersection is $[0:0:1]$. On $Z=1$ the local quotient is $k[x]_{(x)}/(x^2)$, of length 2. Thus $1\cdot 2=2$ is accounted for at a single tangent point.

For the same conic and the line $Z=0$, the only intersection is $[0:1:0]$. On $Y=1$, set $u=X/Y$ and $v=Z/Y$. The equations become $v-u^2=0$ and $v=0$, again giving length 2. A line at infinity can be tangent just like an affine line.

There are three distinct ways a naive counting claim can fail. Counting only affine points misses intersections at infinity. Counting only distinct points misses tangencies and other nonreduced intersections. Allowing a common component makes the intersection positive-dimensional, so the finite-length formula no longer applies. For example, $XY=0$ and $XZ=0$ share the line $X=0$; no sum over finitely many points can describe their whole intersection.

Two conics: one affine intersection and one at infinity

Take

$$F=YZ-X^2, \qquad G=YZ-2X^2$$

over an algebraically closed field of characteristic different from two. The two conics have no common component. Subtracting their equations gives $X^2=0$, so their common point set consists of

$$P=[0:0:1], \qquad Q=[0:1:0].$$

At P, work in $Z=1$. The local intersection algebra is

$$k[x,y]_{(x,y)}/(y-x^2,y-2x^2)\cong k[x]_{(x)}/(x^2),$$

of length two. At Q, use $Y=1$ with coordinates $u=X/Y$, $v=Z/Y$:

$$k[u,v]_{(u,v)}/(v-u^2,v-2u^2)\cong k[u]_{(u)}/(u^2),$$

also of length two. Thus $i_P(F,G)+i_Q(F,G)=2+2=4$, exactly the product of the degrees. Counting only points in $Z=1$ would miss half the total multiplicity.

A line can account for all its intersections at one point

For the conic $F = YZ - X^2$ and the line $G = Y$, the equations imply $X = Y = 0$, so the only projective intersection is $P = [0 : 0 : 1]$. The chart $Z = 1$ gives

$$k[x,y]_{(x,y)}/(y - x^2, y) \cong k[x]/(x^2),$$

of length two. The line is tangent at P. Replacing Y by Z moves the same phenomenon to $[0 : 1 : 0]$: in the chart $Y = 1$, the equations are $v - u^2 = 0$ and $v = 0$. Both calculations exhibit degree two concentrated at a single point. Multiplicity makes Bézout's number stable under such changes of position.

Where the no-common-component condition is used

If $F = XY$ and $G = XZ$, then X is a common factor and the whole line $X = 0$ lies in the intersection. Algebraically multiplication by G on $S/(F)$ is not injective: the nonzero class of Y is killed because $GY = XYZ \in (F)$. Thus the short exact sequence used to compute the Hilbert series fails at its first term. The geometric and algebraic failures are the same obstruction. The theorem cannot be repaired by summing finite point multiplicities along that common line, because the local intersection rings there are not all zero-dimensional.

20.6 Arithmetic genus

For a degree-d plane curve, expand the hypersurface formula:

$$P_C(m) = \binom{m+2}{2} - \binom{m-d+2}{2} = dm + 1 - \frac{(d-1)(d-2)}{2}.$$

For a projective curve define its *arithmetic genus* by $p_a(C) = 1 - P_C(0)$. This is a definition through the Hilbert polynomial; $P_C(0)$ need not equal the actual degree-zero Hilbert function. The formula gives

$$p_a(C) = \frac{(d-1)(d-2)}{2}.$$

Lines and conics have arithmetic genus 0, and plane cubics have arithmetic genus 1.

The arithmetic genus is defined here entirely by a polynomial. Its comparison with the geometry of singularities will be made after the necessary local-ring constructions in Lecture 21.

20.7 Exercises

Exercise 20.1. Compute the Hilbert function and Hilbert polynomial of a projective line and a plane conic. Check small degrees directly.

Exercise 20.2. Let $F, G \in k[X,Y,Z]$ be homogeneous quadratic polynomials with no common factor, and put $R = k[X,Y,Z]/(F,G)$. For the two conics they define, compute the first five values of $h_R(m) = \dim_k R_m$ from the Hilbert series.

Exercise 20.3. Let k be algebraically closed with $\operatorname{char} k \neq 2$. Let $C = V_+(XZ - Y^2)$ and $D = V_+(X^2 - Z^2)$ in $\mathbb{P}^2$. Find all intersections and their multiplicities.

Exercise 20.4. The affine parabola $y = x^2$ and the horizontal line $y = 1$ meet in two points when $\operatorname{char} k \neq 2$. What happens in characteristic 2?

Exercise 20.5 (Count an intersection with multiplicity)**.** In $\mathbb{P}^2_k$, let $C = V(YZ - X^2)$ and $L = V(Y)$. Find all intersection points and compute their multiplicities. Check the answer against the Hilbert polynomial.

Lecture 21
Algebraic curves, singularities, and normalization

Local rings distinguish singular from nonsingular points. The singularities of two curves will be studied by adjoining rational functions that are integral over their coordinate rings. The cusp and node will be worked out completely: their equations, tangent spaces, function fields, integral closures, and fibers of normalization. Reading: [2], Chapter I, §§4–5, and Chapter II, §3 for normalization; [1], Chapter 5 for integral dependence.

21.1 Curves and their function fields

A *curve* is a variety of dimension 1; a *surface* is a variety of dimension 2. An affine curve has coordinate ring a finitely generated reduced integral k-algebra A of Krull dimension 1. Its function field is

$$k(C) = \mathrm{Frac}(A).$$

The dimension theorem gives $\mathrm{trdeg}_k\, k(C) = 1$: one algebraically independent parameter is enough, although the other functions may be algebraic rather than rational expressions in that parameter. A curve is called rational if its function field is isomorphic to $k(t)$. Rationality is a statement about a dense open part of the curve, not a statement that the whole curve is an affine line or a projective line.

For example, on the affine curve $y = x^2$ every function is a polynomial in x, and the coordinate ring is $k[x]$. The curve is isomorphic to $\mathbb{A}^1$. On $xy = 1$, the coordinate ring is $k[x, x^{-1}]$: this is the punctured affine line, with the same function field $k(x)$ but a different set of allowed globally regular functions. Function fields remember birational equivalence; coordinate rings also remember which denominators are allowed everywhere.

At a point p, the local ring

$$\mathcal{O}_{C,p} = A_{\mathfrak{m}_p}$$

allows precisely those denominators that do not vanish at p. Thus the progression

$$A \;\subseteq\; A_{\mathfrak{m}_p} \;\subseteq\; k(C)$$

means: regular everywhere, regular near this point, and rational somewhere on a dense open set. These inclusions should be kept distinct when examining parametrizations. If the maximal ideal of this local ring is generated by a single nonzero element t, that generator is called a *local parameter*. For example, at the point a of the affine line, $x-a$ is a local parameter.

21.2 Tangent space and the local criterion for singularity

For a point p of an affine variety X, let $\mathfrak{m}$ denote the maximal ideal of its local ring $\mathcal{O}_{X,p}$. The quotient $\mathfrak{m}/\mathfrak{m}^2$ is a vector space over the residue field k: if two scalars in the local ring have the same residue, their difference multiplies $\mathfrak{m}$ into $\mathfrak{m}^2$. It is the *cotangent space*; its dual

$$T_pX = \mathrm{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k)$$

is the *Zariski tangent space*. The point is *nonsingular* if its Noetherian local ring is regular, that is, if $\dim_k \mathfrak{m}/\mathfrak{m}^2 = \dim \mathcal{O}_{X,p}$; otherwise it is *singular*. Over our algebraically closed ground field, we also use *smooth* for a variety nonsingular at every point. These definitions apply in projective affine charts because their local rings agree on overlaps. At a nonsingular point of a curve, Nakayama's lemma gives $\mathfrak{m} = (t)$ for an element t; any such generator is called a *local parameter* at p.

Let $p = (a,b)$ lie on an irreducible plane curve $f(x,y) = 0$. Translate coordinates so that p becomes the origin. Write

$$f = f_1 + f_2 + \cdots,$$

where f_j is homogeneous of total degree j in the translated variables. The constant term is zero because p is on the curve. Formal partial derivatives are defined term by term: $\partial(x^iy^j)/\partial x = ix^{i-1}y^j$ and $\partial(x^iy^j)/\partial y = jx^iy^{j-1}$, with the integers interpreted in k. Write f_x, f_y for these derivatives. Modulo $(x,y)^2$, only the linear part survives:

$$f_1 = f_x(p)x + f_y(p)y.$$

In the curve's local ring, the cotangent space is the vector space $\mathfrak{m}_p/\mathfrak{m}_p^2$. It is generated by the classes of x, y, and its relations are exactly the linear relation $f_1 = 0$. To see this last assertion, any multiple hf contributes modulo quadratic terms only $h(p)f_1$; localization by a function nonzero at p merely multiplies this relation by a nonzero scalar.

Definition 21.1. The Zariski tangent space is the dual vector space

$$T_pC = \mathrm{Hom}_k(\mathfrak{m}_p/\mathfrak{m}_p^2, k).$$

A point of a curve is nonsingular if $\dim_k(\mathfrak{m}_p/\mathfrak{m}_p^2) = 1$, and singular otherwise. Equivalently, its one-dimensional Noetherian local ring is regular.

The preceding calculation proves the plane-curve criterion

$$p \text{ nonsingular} \quad \Longleftrightarrow \quad (f_x(p), f_y(p)) \neq (0,0).$$

In the nonsingular case the tangent line consists of vectors (u,v) satisfying $f_x(p)u + f_y(p)v = 0$. If both derivatives vanish, the Zariski tangent space is the whole two-dimensional plane. This does not mean that the curve has become two-dimensional: it means that its first-order equations fail to impose the expected linear restriction. Over $\mathbb{C}$, a nonsingular algebraic curve with its Euclidean topology is locally a real two-dimensional smooth manifold. This analytic comparison follows from the implicit function theorem and is stated here without proof; the algebraic local-ring criterion above is what will be used in our calculations.

For a finer picture, take the first nonzero homogeneous part f_m of f. The equation $f_m = 0$, with its multiplicities retained, defines the tangent cone of the plane curve at the point, and m is the multiplicity of the curve there. Algebraically the tangent cone retains the graded ring $k[x,y]/(f_m)$; its underlying zero set alone forgets repeated factors. Unlike the tangent space, the tangent cone retains this higher-order leading equation and its repeated factors. The word "cone" refers to invariance under scaling the coordinates.

For the cusp $y^2 - x^3 = 0$, the tangent space at the origin is the whole plane, whereas its tangent cone is $y^2 = 0$, a double line. For $y^2 - x^2 - x^3 = 0$ in characteristic different from 2, the tangent cone is

$$y^2 - x^2 = (y-x)(y+x),$$

a pair of distinct lines. This suggests one repeated tangent direction for the cusp and two distinct directions for the node; normalization makes that suggestion precise.

21.3 Integral closure and normalization

Recall that an element of an extension ring is integral over A if it satisfies a monic polynomial with coefficients in A. For a domain A, its integral closure in $\mathrm{Frac}(A)$ is

$$\overline{A} = \{u \in \mathrm{Frac}(A) : u \text{ is integral over } A\}.$$

A domain is *normal* if $\overline{A} = A$. If $\overline{A}$ is finitely generated over k, it is the coordinate ring of an affine variety $\widetilde{C}$. The inclusion $A \hookrightarrow \overline{A}$ then induces a morphism $\nu : \widetilde{C} \to C$, called the *normalization*. It also induces $\mathrm{Spec}\,\overline{A} \to \mathrm{Spec}\,A$ on prime spectra. The classical curves consist of the maximal-ideal points, as explained in Lecture 17; they are not identified with all points of those spectra. A morphism of affine varieties is called *finite* if the induced target-to-source ring map makes the source coordinate ring a finitely generated module over the target coordinate ring.

For a finitely generated integral algebra over a field, the normalization is finite: $\overline{A}$ is a finitely generated A-module. Normalizations of curves over an algebraically closed field are nonsingular. These are general theorems of normalization theory; we state them here without their general proofs, as background for the terminology. See [2], Chapter II, §3 and Chapter I, §6. For the two curves below, finiteness, normality, and nonsingularity will all follow from explicit calculations, so no unproved general assertion is needed to identify their normalizations.

We shall use that $k[t]$ is integrally closed in $k(t)$. Write $u = p/q$ with $\gcd(p,q) = 1$. A monic equation gives

$$u^r + a_1 u^{r-1} + \cdots + a_r = 0, \qquad a_i \in k[t],$$
$$p^r + a_1 p^{r-1} q + \cdots + a_r q^r = 0.$$

Thus $q \mid p^r$. Unique factorization and $\gcd(p,q) = 1$ imply $q \in k^\times$, hence $u \in k[t]$.

21.4 Normalization of the cusp

Consider

$$A = k[x,y]/(y^2 - x^3), \qquad x \longmapsto t^2, \quad y \longmapsto t^3.$$

This identifies A with $k[t^2,t^3] \subset k[t]$. Indeed, division by the monic polynomial $y^2 - x^3$ gives

$$F(x,y) = Q(x,y)(y^2 - x^3) + a(x) + yb(x).$$

If $F(t^2,t^3) = 0$, then

$$a(t^2) + t^3 b(t^2) = 0.$$

The two summands contain only even and only odd powers, respectively. Comparing coefficients gives $a = b = 0$, so the kernel is exactly $(y^2 - x^3)$.

In the fraction field, $t = t^3/t^2 = y/x$, so $\mathrm{Frac}(A) = k(t)$. The element t is integral over A because it satisfies $T^2 - x = 0$. Moreover,

$$k[t] = A + At,$$

since every even power of t belongs to A and every odd power is t times an even power. Thus $k[t]$ is finite over A. It is integrally closed by the argument above. If $u \in k(t)$ is integral over A, its monic equation also has coefficients in $k[t]$, so normality of $k[t]$ forces $u \in k[t]$. Conversely, every element of $k[t]$ is integral over A, since the extension is finite. We have proved

$$\overline{A} = k[t].$$

The normalization map is therefore exactly

$$\nu : \mathbb{A}^1 \longrightarrow C, \qquad t \longmapsto (t^2,t^3).$$

It is an isomorphism away from the origin, with inverse $t = y/x$. Over the origin there is just the point $t = 0$. In fact, the map is bijective on k-points: if $x \neq 0$, $t = y/x$ recovers the unique parameter, and if $x = 0$ the equation forces $y = 0$.

The map is not an isomorphism at the origin. Every element of A has zero coefficient of t. If $t \in A_{(t^2,t^3)}$, then

$$t = \frac{a}{b}, \qquad a,b \in A, \quad b(0) \neq 0, \qquad \text{so } tb = a.$$

Comparison of the coefficients of t gives $b(0) = 0$, a contradiction. Thus

$$t \in \mathrm{Frac}(A), \qquad t \notin A_{(t^2,t^3)}, \qquad t \in k[t]_{(t)}.$$

Normalization supplies the missing local parameter.

21.5 Normalization of the node

Assume for this section that $\mathrm{char}\, k \neq 2$. Let

$$C = V(y^2 - x^2(x+1)).$$

The polynomial is irreducible: as a quadratic in y over $k(x)$, it would factor only if $x^2(x+1)$ were a square, but the factor $x+1$ has odd exponent. Gauss's lemma then gives irreducibility in $k[x,y]$. Thus C is an irreducible curve, even though its singular point has two tangent directions.

In its function field put $t = y/x$. The equation gives $t^2 = x+1$, so

$$x = t^2 - 1, \qquad y = t(t^2-1).$$

Substitution identifies the coordinate ring with

$$A = k[t^2-1, t(t^2-1)] \subset k[t].$$

One can verify the kernel exactly as for the cusp: a remainder $a(x) + yb(x)$ maps to an even polynomial plus an odd polynomial, and these can vanish together only when both are zero. Again $\mathrm{Frac}(A) = k(t)$, and t satisfies the monic equation $T^2 - (x+1) = 0$. Since $k[t] = A + At$, the extension is finite. The same integral-closure argument proves that its normalization is $k[t]$.

The normalization map is

$$\nu(t) = (t^2-1, t(t^2-1)).$$

Outside the origin $x \neq 0$, so $t = y/x$ gives the inverse. Over the origin the equation is $t^2 = 1$, yielding the two distinct points $t = 1$ and $t = -1$. Here a *branch* of the curve at a point means a point of its normalization above that point, together with the local behavior there. Thus normalization separates the two branches at the node. Near $t = 1$, the function $x = (t-1)(t+1)$ is a unit times the local parameter $t-1$, and $y/x = t$ has value 1. Thus the branch has direction $y = x$. Near $t = -1$, the corresponding direction is $y = -x$.

A useful exact description is

$$A = \{h(t) \in k[t] : h(1) = h(-1)\}.$$

Certainly every polynomial in x, y takes equal values at the two points, because both map to the origin. Conversely, divide $h(t)$ by $t^2 - 1$:

$$h(t) = (t^2-1)q(t) + at + b.$$

Equality of the values at 1 and -1 gives $2a = 0$, hence $a = 0$. Write $q(t) = q_0(t^2) + tq_1(t^2)$. Then

$$h(t) = b + xq_0(x+1) + yq_1(x+1) \in A.$$

This makes the identification of the two points completely visible in the ring of functions: precisely those functions that agree at the two points descend to the node.

The cusp and node lose one vector-space dimension of functions when compared with their normalizations, but they lose it in different ways. At the cusp a first-order parameter is missing at one point. At the node two distinct points are required to have the same function value. Neither phenomenon changes the function field. This is why function fields alone do not detect singularities.

The conductor measures which functions already descend

For a finite inclusion of domains $A \subseteq B$ with the same fraction field, the conductor is

$$\mathfrak{c} = \{a \in A : aB \subseteq A\}.$$

It is an ideal of both A and B. Indeed if $a \in \mathfrak{c}$ and $b \in B$, then $ab \in A$ and $(ab)B \subseteq aB \subseteq A$, so $ab \in \mathfrak{c}$. It is also $\mathrm{Ann}_A(B/A)$: a scalar kills every class $b + A$ exactly when $ab \in A$ for every b.

For the cusp, $A = k[t^2, t^3]$ consists of the polynomials with zero coefficient of t. Every power t^n with $n \geq 2$ belongs to A, so

$$B/A = k[t]/k[t^2, t^3] \cong k \cdot [t]$$

as a vector space. A polynomial $a \in A$ satisfies $at \in A$ exactly when its constant term is zero. Because $B = A + At$, checking this single extra generator suffices to test $aB \subseteq A$. Hence

$$\mathfrak{c} = (t^2, t^3) \subseteq A, \qquad \mathfrak{c} = t^2 k[t] \subseteq B.$$

The same ideal has different generating lists in the two rings.

For the node in characteristic different from two, recall

$$A = \{h \in k[t] : h(1) = h(-1)\}.$$

The linear map $k[t] \to k$, $h \mapsto h(1) - h(-1)$, is onto and has kernel A. It therefore identifies B/A with a one-dimensional vector space. If $a \in A$ has common value c at $1, -1$, then at has values $c, -c$. It lies in A exactly when $c = 0$. Again $B = A + At$, so

$$\mathfrak{c} = (t^2-1)k[t] \subseteq B, \qquad \mathfrak{c} = (x, y) \subseteq A.$$

For the last equality, write a multiple $(t^2-1)q(t)$ as

$$xq_0(x+1)+yq_1(x+1), \qquad q(t)=q_0(t^2)+tq_1(t^2).$$

Both quotient spaces B/A have dimension one, but their geometric meaning differs: the cusp is missing one linear coefficient at a single parameter value; the node identifies the values at two parameter points.

The local defect and the genus calculation

For a singular point p of an integral curve, its normalization defect is the finite length

$$\delta_p = \text{length}_{\mathscr{O}_{C,p}}\big((\nu_*\mathscr{O}_{\widetilde{C}})_p/\mathscr{O}_{C,p}\big).$$

For an affine finite normalization $A \subseteq B$, the module in parentheses is $(B/A)_{\mathfrak{m}_p}$. The notation involves the direct image sheaf, defined later; the displayed module provides a purely algebraic interpretation here. At the cusp and node just computed it has length one, and it vanishes away from the singular point. At a node the numerator contains information from both points of the normalization over p, so replacing it by the local ring at only one branch would give the wrong invariant.

For a projective integral curve the normalization sequence is

$$0 \to \mathscr{O}_C \to \nu_*\mathscr{O}_{\widetilde{C}} \to \mathscr{Q} \to 0,$$

where $\mathscr{Q}$ is supported at the singular points and its local lengths are the δ_p. Once the sheaf-cohomology comparison theorems are available, additivity of Euler characteristic gives

$$p_a(C) = g(\widetilde{C}) + \sum_p \delta_p.$$

This last deduction uses the cohomological interpretation $p_a = 1-\chi(\mathscr{O})$ and the cohomology of a finite morphism; it is not being substituted for the earlier elementary Hilbert-polynomial computation. In our examples that computation gives $p_a = 1$, the explicit projective normalization gives $g = 0$, and the local module calculation supplies $\delta = 1$. The three independent calculations agree.

21.6 Arithmetic genus and geometric genus

We use the standard theorem that $p_a(C) = 1 - P_C(0)$ is independent of the projective embedding of a projective curve. This invariance is a comparison theorem, stated here without proof; see [2], Chapter III. It allows the following definition to depend only on the curve.

For an integral projective curve C, its *geometric genus* is the arithmetic genus of its nonsingular projective normalization $\widetilde{C}$:

$$g(C) = p_a(\widetilde{C}).$$

The normalization is obtained by taking integral closures on affine charts and identifying the resulting rings on their common open subsets; integral closure commutes with localization. We use here the general fact that normalization of a projective curve is again projective. For nonsingular C, all local rings are integrally closed, so normalization makes no change and $g(C) = p_a(C)$. Thus arithmetic and geometric genus agree for nonsingular projective curves. Their cohomological interpretation will be discussed after cohomology of functions on open sets has been defined; see [2], Chapters III–IV.

The projective cuspidal cubic $Y^2Z = X^3$ is parametrized by

$$[s:t] \longmapsto [s^2t : s^3 : t^3].$$

On $t \neq 0$ this is the affine cusp parametrization with parameter s/t. The only added point is $[0:1:0]$; on $Y = 1$ its equation is $Z = X^3$, a nonsingular affine chart. For the nodal cubic $Y^2Z = X^2(X+Z)$, with $\operatorname{char} k \neq 2$, the parametrization is

$$[s:t] \longmapsto [t(s^2-t^2) : s(s^2-t^2) : t^3].$$

The coordinates cannot all vanish. On $t \neq 0$ it is the affine normalization already calculated. The only additional point is again $[0:1:0]$, which is nonsingular: the derivative in Z of $Y^2Z - X^2(X+Z)$ has value 1 there. More explicitly, put $s = 1$ near the added point of the parameter line. On $Y = 1$ the nodal map has coordinates $X = t$ and $Z = t^3/(1-t^2)$, with inverse $t = X$ near zero. The cusp map has coordinates $X = t$ and $Z = t^3$, again with inverse $t = X$. Thus the displayed maps are isomorphisms near the added point and extend the affine normalizations by one nonsingular point. Both projective normalizations are consequently $\mathbb{P}^1$. Their geometric genera are therefore 0, whereas their arithmetic genera are 1 by their cubic equations. A singular plane equation need not have the same arithmetic genus as its nonsingular normalization.

21.7 Exercises

Exercise 21.1. For $C = V(y-x^2)$ and $p = (a,a^2)$, describe $\mathcal{O}_{C,p}$ and its maximal ideal. Is $1/(x-a)$ regular at p? Is $1/(x-b)$ regular there when $b \neq a$?

Exercise 21.2. Locate all singular points of the affine cusp $y^2 = x^3$. Do not assume characteristic zero.

Exercise 21.3. Put $A = k[t^2,t^3] \subset k[t]$. Describe $k[t]/A$ as a k-vector space. Then find the conductor

$$\mathfrak{c} = \{a \in A : a\,k[t] \subseteq A\}.$$

Exercise 21.4. Assume $\operatorname{char} k \neq 2$. Let C be the nodal curve $y^2 = x^2(x+1)$, with coordinate ring $A = k[t^2-1, t(t^2-1)] = \{h \in k[t] : h(1) = h(-1)\}$. Show that the origin is its unique singular point, and use this description to compute $\dim_k(k[t]/A)$.

Exercise 21.5 (A bijective morphism which is not an isomorphism)**.** For the cusp $C = V(y^2 - x^3) \subseteq \mathbb{A}^2_k$ over an algebraically closed field of characteristic zero, consider $\nu : \mathbb{A}^1 \to C$, $t \mapsto (t^2, t^3)$. Prove that it is a bijection on points and is birational, but is not an isomorphism.

Lecture 22
Blowing up a point: complete calculations in two charts

Blowing up the origin of the affine plane replaces it by the set of directions through that point. Explicit affine coordinates describe the resulting surface and the changes to curves on it. We will distinguish total and strict transforms and recover the normalizations of the cusp and node from Lecture 21. Reading: [2], Chapter I, §4 and the curve-singularity exercises in §5; Chapter II, §7 for the general construction.

22.1 The incidence construction

Away from the origin, the point (x,y) determines a direction $[x:y] \in \mathbb{P}^1$. The rational map

$$\mathbb{A}^2 \dashrightarrow \mathbb{P}^1, \qquad (x,y) \longmapsto [x:y]$$

is undefined at the origin. Instead of trying to choose one direction there, we construct a space whose points remember a direction as additional data.

Definition 22.1. The blowup of $\mathbb{A}^2$ at the origin is

$$B = \{(x,y,[u:v]) \in \mathbb{A}^2 \times \mathbb{P}^1 : xv = yu\}.$$

Its projection is $\pi : B \to \mathbb{A}^2$, $(x,y,[u:v]) \mapsto (x,y)$.

The equation is homogeneous in u,v, so it is well defined on the projective factor. If $(x,y) \neq (0,0)$, it forces $[u:v] = [x:y]$. Therefore the blowup agrees with the graph of the direction map outside the origin. At $(0,0)$, the equation imposes no condition on $[u:v]$, so

$$E = \pi^{-1}(0) \cong \mathbb{P}^1.$$

This is the *exceptional curve*. Its points are the lines through the original origin, including the vertical line. We now show explicitly that this set-theoretic description comes from well behaved algebraic charts.

22.2 The two affine charts and their transition

On $u \neq 0$, put $t = v/u$. Then

$$xv = yu \iff y = xt, \qquad k[x,y,t]/(y - xt) \cong k[x,t].$$

The projection is

$$\pi_x : \mathbb{A}^2_{x,t} \longrightarrow \mathbb{A}^2_{x,y}, \qquad (x,t) \longmapsto (x,xt).$$

The exceptional curve here is $x = 0$, with coordinate t recording the slope.

On $v \neq 0$, put $s = u/v$. Similarly,

$$xv = yu \iff x = sy, \qquad k[x,y,s]/(x - sy) \cong k[s,y],$$

with projection

$$\pi_y : \mathbb{A}^2_{s,y} \longrightarrow \mathbb{A}^2_{x,y}, \qquad (s,y) \longmapsto (sy,y).$$

Its exceptional curve is $y = 0$. This chart includes the vertical direction $s = 0$, which the slope t chart misses.

On the overlap, $t \neq 0$ and $s \neq 0$, and the transition formulas are

$$s = t^{-1}, \qquad y = xt; \qquad t = s^{-1}, \qquad x = sy.$$

All denominators are invertible on their stated domains. Restricting to the exceptional curve gives precisely the transition $s = t^{-1}$ for $\mathbb{P}^1$. Since both surface charts are affine planes, the blowup is nonsingular. We did not introduce a new singularity by inserting the exceptional curve.

Every exceptional point lies in the closure of the graph away from the origin. In the first chart, fix $t = a$ and let x vary: the line (x,a) contains its exceptional point $(0,a)$ in the closure of $x \neq 0$. The other chart gives the same argument for the vertical direction. Hence B is the closure of the graph, and it is irreducible because that graph is isomorphic to the irreducible punctured plane and is dense.

Proposition 22.2. *The projection π restricts to an isomorphism*

$$B \setminus E \xrightarrow{\sim} \mathbb{A}^2 \setminus \{0\}.$$

Proof. The inverse is $\iota(x,y) = (x,y,[x:y])$, with regular chart expressions

$$\iota|_{x\neq 0}(x,y) = (x,y/x), \qquad \iota|_{y\neq 0}(x,y) = (x/y,y).$$

On the overlap, $s = x/y = (y/x)^{-1} = t^{-1}$, so they agree. Direct substitution gives

$$\pi \circ \iota = \mathrm{id}_{\mathbb{A}^2\setminus\{0\}}, \qquad \iota \circ \pi = \mathrm{id}_{B\setminus E}.$$

□

22.3 Total and strict transforms

Let $C \subset \mathbb{A}^2$ be a curve through the origin, defined by $f(x,y) = 0$. The total transform is defined by the pulled-back equation $f \circ \pi = 0$. It retains factors that describe the exceptional curve, including their multiplicities. The strict transform is the closure

$$\widetilde{C} = \overline{\pi^{-1}(C \setminus \{0\})} \subset B.$$

This closure follows the original curve away from the center and then adds its limiting points over the center. It does not automatically contain the whole exceptional curve.

Write $f = f_m + f_{m+1} + \cdots$, where $f_m \neq 0$ is the first homogeneous part. In the first chart,

$$f(x,xt) = x^m\big(f_m(1,t) + x f_{m+1}(1,t) + \cdots\big).$$

The total transform has the factor x^m. After restricting to $x \neq 0$, that factor is invertible, so the equation becomes

$$g(x,t) = f_m(1,t) + x f_{m+1}(1,t) + \cdots = 0.$$

Since $g(0,t) = f_m(1,t) \neq 0$, we have $x \nmid g$. Unique factorization gives

$$g \mid x^r h \implies g \mid h, \qquad (g) : x^\infty = (g).$$

Hence

$$(f(x,xt)) : x^\infty = (x^m g) : x^\infty = (g),$$

so $g = 0$ is the strict transform in this chart. This is the localization-and-contraction principle of Lecture 19.

On E, the strict-transform equation reduces to $f_m(1,t) = 0$. In the second chart it reduces to $f_m(s,1) = 0$. Thus the points at which the strict transform meets the exceptional curve are the projective tangent directions of the original curve. Repeated factors of f_m can produce nontransverse intersections with E; one should not infer smoothness or transversality merely from the number of distinct directions.

22.4 Blowing up the cusp

For $C : y^2 = x^3$, the first chart gives

$$(xt)^2 - x^3 = x^2(t^2 - x).$$

The total transform has exceptional factor x^2, while the strict transform is

$$\widetilde{C} : \quad t^2 - x = 0.$$

It is nonsingular: its derivative with respect to x is -1. Its coordinate ring is $k[x,t]/(t^2-x) \cong k[t]$, and the projection to the original plane becomes

$$x = t^2, \qquad y = xt = t^3.$$

This is exactly the normalization already proved in Lecture 21.

In the second chart, substituting $x = sy$ gives

$$y^2 - (sy)^3 = y^2(1 - s^3 y).$$

The strict transform has equation $1 - s^3 y = 0$. It does not meet $E = (y = 0)$. On this strict transform s is invertible, so this whole piece is already in the overlap with the first chart. Hence the first chart contains the entire strict transform of the affine cusp, not merely a neighborhood of its exceptional point.

The strict transform meets E at the single point $x = t = 0$, but the intersection is not transverse. The local intersection quotient is

$$k[x,t]_{(x,t)}/(x, t^2 - x) \cong k[t]_{(t)}/(t^2),$$

which has basis $1, t$ and length 2. Moreover,

$$T_0\widetilde{C} = V(x) = T_0 E, \qquad i_0(\widetilde{C}, E) = 2.$$

The strict transform is nonsingular but tangent to E; see Figure 22.1.

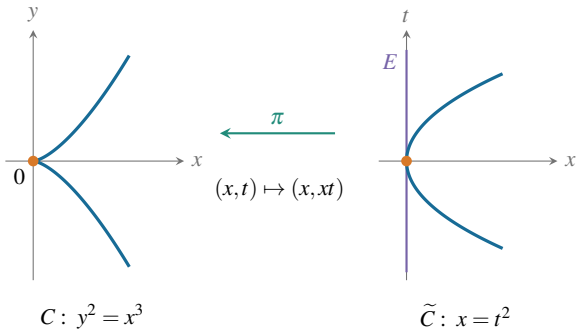


Fig. 22.1 Real loci of the cusp and its strict transform in the $u \neq 0$ chart. The exceptional curve is purple. The smooth strict transform meets it at one point with intersection multiplicity 2.

22.5 Blowing up the node

Assume again that $\operatorname{char} k \neq 2$, and take $C : y^2 = x^2(x+1)$. In the first chart,

$$(xt)^2 - x^2(x+1) = x^2(t^2 - x - 1).$$

The strict transform is $t^2 - x - 1 = 0$, so its coordinate ring is $k[t]$, with

$$x = t^2 - 1, \qquad y = t(t^2 - 1).$$

The derivative in x is -1, so the strict transform is smooth. It meets $E = (x = 0)$ at $t = 1$ and $t = -1$, precisely the two points of the normalization over the node.

The other chart gives

$$y^2 - (sy)^2(sy+1) = y^2(1 - s^2 - s^3 y).$$

The strict-transform equation $1 - s^2 - s^3 y = 0$ has no solution with $s = 0$. Hence this chart also contributes only points in the overlap with the first. The full strict transform of the affine nodal curve is therefore the affine line already parametrized by t.

At either exceptional point, the equation $t^2 - 1$ has a simple root because its derivative $2t$ is nonzero. Therefore

$$\operatorname{length}\bigl(k[x,t]_{(x,t\mp 1)}/(x, t^2 - x - 1)\bigr) = 1.$$

The two branches have become two transverse intersections with the exceptional curve. The contrast with the cusp is visible in the lowest homogeneous term: $y^2 - x^2$ has two distinct projective roots, whereas y^2 has one double root.

Why the exceptional factor is removed in the strict transform

On the chart $y = xt$, the pullback of a plane equation $f(x,y) = 0$ is $f(x,xt) = 0$. If the smallest total degree of a nonzero term of f is m, then

$$f(x,xt) = x^m\bigl(f_m(1,t) + x f_{m+1}(1,t) + \cdots\bigr).$$

The factor x^m accounts for the exceptional divisor with its multiplicity in the total transform. The strict transform is the closure of the part over $x \neq 0$, so its ideal is obtained by saturation by x. For a plane hypersurface, removing the maximal common power of x from its equation gives this saturated principal ideal. To see this algebraically, the remaining polynomial is not divisible by the prime element x; unique factorization shows that $x^r h$ being divisible by it implies that h is divisible by it.

For the cusp $y^2 = x^3$, the total transform is $x^2(t^2 - x) = 0$ and the strict transform is $x = t^2$. The parameter t gives

$$(x,y) = (t^2, t^3).$$

The strict transform is smooth, but it meets the exceptional divisor $x = 0$ with multiplicity two, since its local quotient by x is $k[t]_{(t)}/(t^2)$. Thus one blow-up makes the strict transform nonsingular, while the union of that transform with the exceptional divisor is still tangent. Resolving a curve as a variety and making the whole total transform have transverse components are different requirements.

For the node $y^2 = x^2 + x^3$, the strict transform is $t^2 = 1 + x$. It meets $x = 0$ at $t = 1, -1$. Near $t = 1$, the factor $t + 1$ is a unit and $x = (t-1)(t+1)$ vanishes to order one. The same holds at $t = -1$. Thus each intersection with the exceptional divisor is transverse. The characteristic restriction $2 \neq 0$ is used when making $t + 1$ or $t - 1$ a unit.

The exceptional curve has a nontrivial normal bundle

Use the two charts

$$U_x = \operatorname{Spec} k[x,t], \quad y = xt, \qquad U_y = \operatorname{Spec} k[s,y], \quad x = sy.$$

On their overlap, $s = t^{-1}$ and $y = tx$. The exceptional curve E is $x = 0$ on U_x and $y = 0$ on U_y, with coordinate transition $s = t^{-1}$, so $E \cong \mathbb{P}^1$.

Its ideal modulo its square is locally generated by the classes $\bar{x}$ and $\bar{y}$. Their transition is

$$\bar{y} = t\bar{x}.$$

Under the convention for $\mathscr{O}(n)$ in Lecture 26, the conormal line bundle $\mathscr{I}_E/\mathscr{I}_E^2$ is therefore $\mathscr{O}_E(1)$. Dualizing gives the normal line bundle $\mathscr{O}_E(-1)$. This calculation uses only the chart transition; the line-bundle terminology and its gluing interpretation will be developed in Lecture 26. It explains the minus-one normal direction associated with the exceptional curve.

22.6 The Rees algebra and relative projective construction

The construction can be expressed using the ideal of the center. Put $A = k[x,y]$ and $I = (x,y)$. Introduce an auxiliary symbol T and define the graded Rees algebra

$$\mathscr{R}(I) = A \oplus IT \oplus I^2T^2 \oplus \cdots \subset A[T].$$

The coefficient ring A is placed in degree 0, and xT, yT have degree 1. This grading records how many factors from the ideal I occur; it is different from grading x and y themselves by polynomial degree.

There is a surjective graded homomorphism

$$A[U,V] \longrightarrow \mathscr{R}(I), \qquad U \longmapsto xT, \quad V \longmapsto yT.$$

Its kernel is $(xV - yU)$. To check this without guessing, first invert x: then $V = (y/x)U$, and the kernel is generated by that relation. Globally the polynomial $xV - yU$ is irreducible in the UFD $k[x,y,U,V]$, so its ideal is prime and does not contain x. Its contraction from the localization at x is therefore itself. Any relation in the original kernel becomes this relation after localization and consequently already belongs to $(xV - yU)$.

We next define the relative projective construction used in this description. Let $S = \bigoplus_{d\geq 0} S_d$ be a graded ring with $S_0 = A$, generated over A by finitely many degree-one elements $z_0, \ldots, z_r$. The set $\operatorname{Proj}_A S$ consists of homogeneous prime ideals that do not contain the ideal $S_+ = \bigoplus_{d>0} S_d$. Closed sets are the subsets of such primes containing a specified homogeneous ideal. The open chart $D_+(z_i)$ consists of the primes not containing z_i and is identified with

$$\operatorname{Spec}\bigl(S[z_i^{-1}]\bigr)_0.$$

To see the identification, localization first replaces such primes by homogeneous primes of $S[z_i^{-1}]$. Since z_i is an invertible degree-one element, every homogeneous fraction is a degree-zero fraction times a power of z_i. Hence a homogeneous prime is determined by its degree-zero part, and every prime of that degree-zero ring gives a homogeneous prime. On an overlap, the two chart rings are localized further by z_j/z_i and z_i/z_j and become the same ring. The charts are therefore identified on these common open subsets. This chart description also specifies their rings of regular functions; the general language for this structure is developed in Lecture 25.

Apply the construction to the Rees algebra. Its relation $xV = yU$ gives the same two chart rings as the blowup. In particular, its closed points over k are the previously described incidence points $((x,y),[u:v])$ satisfying $xv = yu$; the prime spectrum description also retains nonclosed points. The chart rings are

$$\bigl(\mathscr{R}(I)[(xT)^{-1}]\bigr)_0 = A[y/x], \qquad \bigl(\mathscr{R}(I)[(yT)^{-1}]\bigr)_0 = A[x/y].$$

These are exactly $k[x,t]$ with $y = xt$, and $k[s,y]$ with $x = sy$.

For a ring A and a finitely generated ideal I, the same expression $\operatorname{Proj}_A\bigl(\bigoplus_{n\geq 0} I^n T^n\bigr)$, with the affine chart rings just specified, defines the blowup along I. The ideal specifies the *center* of the blowup, whose underlying closed set is $V(I)$. The general characterization of this construction by a universal property is beyond these computations; see [2], Chapter II, §7. Here the two charts already prove the geometric assertions for the plane.

Factoring a map through the blow-up

Here is the concrete universal property behind the incidence equations. Suppose a morphism $f : Y \to \mathbb{A}^2$ pulls the ideal (x,y) back to an invertible ideal on Y. On a neighborhood where this ideal is generated by a non-zero-divisor a, write

$$f^*x = au, \qquad f^*y = av.$$

Because a generates the pulled-back ideal, $a = rau + sav$ for suitable local functions r,s. Cancelling a gives $ru + sv = 1$. Hence u,v never vanish simultaneously and define a map to $\mathbb{P}^1$. The formula

$$z \longmapsto \big(f(z), [u(z) : v(z)]\big)$$

lands in the blow-up because $f^*x\,v = f^*y\,u$. If the generator is changed to $a' = ca$ with c a unit, then $(u',v') = (c^{-1}u, c^{-1}v)$, so the projective direction does not change. The local lifts therefore glue.

The lift is unique. On a neighborhood where u is a unit, any direction $[S : T]$ satisfying the incidence equation must obey $uT = vS$ after cancelling a. It follows that S is nonvanishing there and $T/S = v/u$, so the direction is forced to be $[u : v]$. The same argument on the neighborhood where v is a unit completes the proof. The cancellation hypothesis is the reason the property is formulated for an invertible ideal, rather than for an arbitrary ideal generated by a single zero divisor.

22.7 Exercises

Exercise 22.1. Let $\pi : B = \mathrm{Bl}_0 \mathbb{A}^2_k \to \mathbb{A}^2_k$ be the blow-up at the origin, with exceptional curve $E = \pi^{-1}(0)$. For $a \in k$, find the closure in B of the inverse image of the punctured line $y = ax$, and determine where it meets E. Repeat for $x = 0$.

Exercise 22.2. Blow up $\mathbb{A}^2_k$ at the origin. For the line $y = 0$, compute the total and strict transforms in the charts $y = xt$ and $x = sy$. Why does one chart have no part of the strict transform?

Exercise 22.3. Why is it incorrect to say that blowing up $\mathbb{A}^2_k$ at the origin on the cusp $y^2 = x^3$ replaces its singular point by an entire projective line on the strict transform?

Exercise 22.4. Blow up $\mathbb{A}^2_k$ at the origin, and let E be the exceptional curve. For the reducible curve $xy = 0$, describe the strict transforms of its two components and their intersections with E.

Exercise 22.5. Let A be a domain and $0 \neq f \in A$. Describe the blow-up of $\mathrm{Spec}\,A$ along the principal ideal $I = (f)$, and prove that its projection to $\mathrm{Spec}\,A$ is an isomorphism.

Exercise 22.6 (Affine charts of a blow-up)**.** Blow up $\mathbb{A}^2_k$ at the origin. Write its two standard affine charts and transition formula. For the cusp $y^2 = x^3$, compute its strict transform in the chart $y = xv$, and compare it with the total transform.

Lecture 23
Elliptic curves and the chord–tangent group law

A nonsingular cubic with a specified rational point carries an abelian group law. We derive its addition formulas from line intersections, including the tangent and vertical cases. The associativity theorem uses an additional result from the theory of divisors; that input is stated explicitly below. References are [2, Chapter IV, §4] and, for divisors, [2, Chapter II, §6].

23.1 The equation and its point at infinity

Throughout this lecture, k is a field of characteristic different from 2 and 3, and $\overline{k}$ is an algebraic closure of k. Examples are $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and finite fields of characteristic at least 5. Extending the ground field means viewing the same defining equations with coefficients in the larger field; it allows additional solutions and may reveal components or singular points that were not visible over k.

We recall the terms needed in the definition. A curve has dimension one, and a projective curve is a closed subvariety of some projective space. A curve is *geometrically integral* if after extension to $\overline{k}$ it is irreducible and reduced. Here reduced means that its coordinate rings have no nonzero nilpotent elements. A curve is *smooth over k* if it is nonsingular after this extension. For a plane curve, nonsingularity is the partial-derivative criterion from Lecture 21. For a smooth projective geometrically integral curve, its genus is $g = p_a(C)$, where $p_a(C) = 1 - P_C(0)$ is defined by the Hilbert polynomial in Lecture 20. This agrees with the definition through nonsingular normalization in Lecture 21, since a nonsingular curve is already normal. In particular, a smooth plane curve of degree d has genus $(d-1)(d-2)/2$.

A *k-rational point* of a projective curve is a solution represented by homogeneous coordinates in k. The coordinates are not all zero, and multiplying all of them by the same element of $k^{\times}$ does not change the point. We write $C(k)$ for the set of these points.

Definition 23.1. An *elliptic curve over k* is a smooth projective geometrically integral curve of genus one, together with a specified k-rational point O.

In characteristic different from 2 and 3, such a curve admits a *short Weierstrass equation*

$$E: \quad Y^2Z = X^3 + aXZ^2 + bZ^3, \qquad a,b \in k, \qquad 4a^3 + 27b^2 \neq 0,$$

with $O = [0:1:0]$. The existence of this form is a theorem of the theory of curves, which we state without proof; see [2, Chapter IV, §4]. Our computations concern the displayed model. Its degree is three, so the genus formula gives genus one once nonsingularity is checked. A nonsingular plane cubic over an algebraically closed field is irreducible: a decomposition into components would force two positive-degree components to meet, and a point of intersection would be singular. This uses the plane intersection theorem from Lecture 20.

In the chart $Z \neq 0$, put $x = X/Z$ and $y = Y/Z$. The equation becomes

$$y^2 = x^3 + ax + b.$$

At $Z = 0$ the homogeneous equation forces $X = 0$, so O is the only missing point. This projective point will be the identity for the group law.

Proposition 23.2. *The projective cubic above is nonsingular if and only if* $4a^3 + 27b^2 \neq 0$.

Proof. Write $f = y^2 - x^3 - ax - b$. At an affine singular point both partial derivatives must vanish:

$$f_y = 2y = 0, \qquad f_x = -3x^2 - a = 0.$$

Thus $y = 0$ and $x = r$ is a common root of $g(x) = x^3 + ax + b$ and $g'(x) = 3x^2 + a$. Hence

$$a = -3r^2, \qquad b = 2r^3, \qquad 4a^3 + 27b^2 = -108r^6 + 108r^6 = 0.$$

Conversely, suppose this expression is zero. If $a = 0$, then $b = 0$ and $r = 0$ is a common root. If $a \neq 0$, set $r = -3b/(2a)$. Substitution using $27b^2 = -4a^3$ gives $r^2 = -a/3$ and $r^3 = b/2$, hence $g(r) = g'(r) = 0$. This produces a singular affine point. Finally put

$$F = Y^2Z - X^3 - aXZ^2 - bZ^3.$$

At O we have $F_Z = Y^2 - 2aXZ - 3bZ^2 = 1$. Therefore the point at infinity is always nonsingular, and the affine test is sufficient. □

The conventional discriminant is $\Delta = -16(4a^3 + 27b^2)$. Only its nonvanishing matters here. The equation $y^2 = x^3$ fails the test: its cusp is precisely why it does not define an elliptic curve.

23.2 Secant and tangent formulas

Reflection in the x-axis preserves the equation. Define

$$-(x,y) = (x,-y), \qquad -O = O.$$

For two distinct affine points $P=(x_1,y_1)$ and $Q=(x_2,y_2)$ with $x_1\neq x_2$, draw the line through them. It meets the cubic in one further point R, counted with intersection multiplicity. Define $P+Q=-R$. Substitution into the equation gives its coordinates as follows.

The line is $y=mx+c$, where

$$m=\frac{y_2-y_1}{x_2-x_1},\qquad c=y_1-mx_1.$$

Substituting in the cubic gives the monic polynomial

$$x^3-m^2x^2+(a-2mc)x+(b-c^2)=0.$$

Its roots, with multiplicity, are x_1,x_2,x_R. Comparing the coefficient of x^2 with the expansion of $(x-x_1)(x-x_2)(x-x_R)$ gives

$$x_R=m^2-x_1-x_2.$$

Reflection changes y_R to $-y_R$, so the answer is

$$\boxed{x_3=m^2-x_1-x_2,\qquad y_3=m(x_1-x_3)-y_1.} \tag{23.1}$$

All coefficients used in this calculation lie in k. Therefore the third point and its reflection are again k-rational.

Example 23.3. On $y^2=x^3+1$, take $P=(0,1)$ and $Q=(2,3)$. Then

$$\begin{aligned} m&=\frac{3-1}{2-0}=1,\\ x_3&=1^2-0-2=-1,\\ y_3&=1(0-(-1))-1=0. \end{aligned}\qquad\Longrightarrow\qquad P+Q=(-1,0).$$

The intersection calculation verifies all three points:

$$(x+1)^2=x^3+1\quad\Longleftrightarrow\quad x(x-2)(x+1)=0.$$

When $P=Q$, write $2P=P+P$ and use the tangent line. Implicit differentiation of $y^2=x^3+ax+b$ gives $2y\,dy/dx=3x^2+a$. If $y_1\neq 0$, its slope is

$$m=\frac{3x_1^2+a}{2y_1}.$$

The same substitution now has the root x_1 twice. Therefore

$$x(2P)=m^2-2x_1,\qquad y(2P)=m\big(x_1-x(2P)\big)-y_1.$$

The double root records the tangency. Figure 23.1 illustrates the secant construction, including the final reflection.

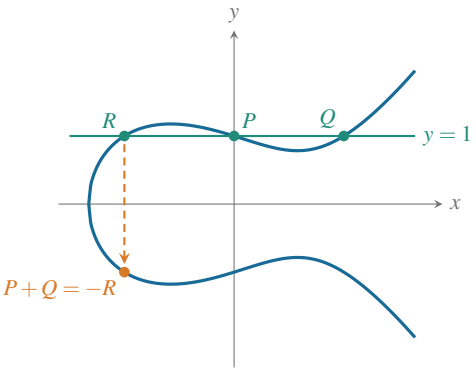


Fig. 23.1 On $y^2 = x^3 - x + 1$, the line through $P = (0,1)$ and $Q = (1,1)$ has third intersection $R = (-1,1)$. Reflection gives $P + Q = (-1,-1)$. The real drawing illustrates the algebraic formulas.

23.3 Vertical lines, the identity, and torsion

A vertical line $x = x_1$ meets the curve at $P = (x_1, y_1)$, $-P = (x_1, -y_1)$, and O. Consequently we define $P + (-P) = O$. If $y_1 = 0$, then $P = -P$, and the tangent is vertical. In this case $2P = O$; we must not divide by $2y_1$.

The complete rule is

$$P + O = O + P = P, \qquad P + Q = \begin{cases} O, & x_1 = x_2,\ y_1 = -y_2, \\ (m^2 - x_1 - x_2,\ m(x_1 - x_3) - y_1), & \text{otherwise}, \end{cases}$$

where $x_3 = m^2 - x_1 - x_2$ and

$$m = \begin{cases} (3x_1^2 + a)/(2y_1), & P = Q,\ y_1 \neq 0, \\ (y_2 - y_1)/(x_2 - x_1), & x_1 \neq x_2. \end{cases}$$

These cases exhaust the possibilities, since

$$x_1 = x_2 \implies (y_1 - y_2)(y_1 + y_2) = 0 \implies y_2 = \pm y_1.$$

The identity rule also has a geometric explanation. The line from O to an affine P is vertical, and its third intersection is $-P$, so reflecting gives P. The tangent at O is $Z = 0$. Its intersection with the cubic is $X^3 = 0$, so the intersection multiplicity at O is three. Such a point is called a *flex* or *inflection point*.

Theorem 23.4 (Group law)**.** *The construction above makes $E(k)$ an abelian group with identity O. In particular,*

$$(P + Q) + R = P + (Q + R)$$

for all points, including tangent and vertical configurations. Addition and inversion are morphisms of algebraic varieties.

This theorem is stated without a full proof. The formulas establish closure, the identity rule, inverses, and commutativity. The additional input for associativity is described in the next section. We may therefore use the language of abelian groups in the following examples. For an integer $n > 0$, nP means the sum of n copies of P; also $0P = O$ and $(-n)P = -(nP)$. A point is a *torsion point* if $nP = O$ for some $n > 0$. Its *order* is the least such positive integer. The identity has order one.

A complete addition computation

Consider $E : y^2 = x^3 - x + 1$ over $\mathbb{Q}$, with $P = (0,1)$ and $Q = (1,1)$. The chord is $y = 1$. Substitution gives

$$1 = x^3 - x + 1 \quad \Longleftrightarrow \quad x(x-1)(x+1) = 0,$$

so its third intersection is $R = (-1,1)$ and $P + Q = (-1,-1)$. The sign change is essential; the third intersection itself is the negative of the sum.

For doubling at P, the tangent slope is

$$m = \frac{3x(P)^2 - 1}{2y(P)} = -\frac{1}{2}.$$

The formulas give

$$x(2P) = m^2 - 2x(P) = \frac{1}{4}, \qquad y(2P) = m\big(x(P) - x(2P)\big) - y(P) = -\frac{7}{8}.$$

Check the equation independently:

$$\left(-\frac{7}{8}\right)^2 = \frac{49}{64} = \left(\frac{1}{4}\right)^3 - \frac{1}{4} + 1.$$

This confirms that the computed sum is on the curve, although it does not by itself prove the group axioms.

On $E' : y^2 = x^3 - x$, the points $(0,0), (1,0), (-1,0)$ have vertical tangent lines. The denominator $2y$ in the tangent-slope formula is zero there. The geometric rule instead gives $2P = O$, consistently with $P = -P$. Formulas with denominators should always be accompanied by their exceptional cases.

Why the cubic must be nonsingular

For $F(X,Y,Z) = Y^2Z - X^3 - aXZ^2 - bZ^3$ in characteristic different from two and three, an affine singular point would satisfy

$$y = 0, \qquad 3x^2 + a = 0, \qquad x^3 + ax + b = 0.$$

Thus it is a repeated root of $x^3 + ax + b$. The condition $4a^3 + 27b^2 \neq 0$ excludes exactly that possibility. At $O = [0:1:0]$, one has $F_Z(O) = 1$, so the point at infinity is nonsingular. The discriminant condition is therefore a geometric smoothness condition, not just a convenient denominator restriction in the addition formulas.

23.4 Divisors and associativity

Symmetry of the secant construction proves commutativity. The identity and inverse rules are already established. Associativity, however, compares two different sequences of line constructions. A sketch of a few lines is not a proof for every configuration.

We explain the divisor-theoretic input in Theorem 23.4, following [2, Chapter IV, §4]. Work over $\overline{k}$. At a nonsingular closed point P of a curve, the local ring is a one-dimensional regular local ring, hence a discrete valuation ring by Lecture 10. A generator t of its maximal ideal is called a *local parameter*. Every nonzero rational function has the form $t^n u$ in the fraction field of this ring, where $n \in \mathbb{Z}$ and u is a unit. Define $\operatorname{ord}_P(f) = n$. Positive order means a zero, negative order means a pole, and order zero means that f is regular and nonzero at P.

A *divisor* is a finite formal integer sum $D = \sum_P n_P [P]$ over closed points P. Its degree is $\deg D = \sum_P n_P$, and it is *effective* if every coefficient n_P is nonnegative. Formal sums are added coefficient by coefficient. A nonzero rational function has a *principal divisor*

$$\operatorname{div}(f) = \sum_P \operatorname{ord}_P(f)[P].$$

Only finitely many terms are nonzero. The finiteness and the identity $\deg \operatorname{div}(f) = 0$ are standard divisor theorems, used here without proof; see [2, Chapter II, §6 and Chapter IV, §1]. For instance, on the projective line $\operatorname{div}(t - a) = [a] - [\infty]$: the function has one simple zero and one simple pole.

Two divisors D, D' are *linearly equivalent*, written $D \sim D'$, if $D - D'$ is principal. Principal divisors form a subgroup, since orders add under multiplication. The quotient of the group of degree-zero divisors by this subgroup is denoted $\operatorname{Pic}^0(E)$. Thus its elements are divisor classes and its addition is inherited from addition of formal sums.

The special case of the Riemann–Roch theorem needed here can now be stated. For a divisor D on a smooth projective genus-one curve, put

$$L(D) = \{0\} \cup \{f \in \overline{k}(E)^\times : \operatorname{div}(f) + D \text{ is effective}\}.$$

This is a vector space over $\overline{k}$: the valuation inequality ensures that adding two functions does not create a worse pole, and multiplication by a nonzero constant leaves orders unchanged. If $D = [O]$, it consists of rational functions regular away from O with at most a simple pole at O, together with zero. The genus-one Riemann–Roch theorem asserts

$$\dim_{\overline{k}} L(D) = \deg D \qquad (\deg D > 0).$$

We use this form of Riemann–Roch without proof; see [2, Chapter IV, §1]. It implies that

$$E(\bar{k}) \longrightarrow \mathrm{Pic}^0(E), \qquad P \longmapsto \text{the class of } [P]-[O],$$

is a bijection. Indeed, given a degree-zero divisor D, the space $L(D+[O])$ has dimension one. Choose a nonzero member f. The effective divisor $\mathrm{div}(f)+D+[O]$ has degree one and is therefore $[P]$ for one point P. This proves surjectivity. If $[P]-[Q]$ is principal, its defining function belongs to $L([Q])$. This space has dimension one and contains all constants, so the function is constant, its divisor is zero, and $P=Q$. This proves injectivity. Transporting addition across the bijection gives an associative operation, since addition of divisor classes is associative.

The operation agrees with the line construction. Let $L(X,Y,Z)=0$ be a line other than $Z=0$ meeting the cubic at P,Q,R, with multiplicity. The quotient L/Z is a rational function on the curve: numerator and denominator have the same homogeneous degree. Its divisor is

$$\mathrm{div}(L/Z) = [P]+[Q]+[R]-3[O].$$

The pole term comes from the triple intersection with $Z=0$. Consequently the intersection divisor is linearly equivalent to $3[O]$. Thus

$$[P]+[Q]+[R] \sim 3[O].$$

For a vertical line, $[R]+[-R]+[O] \sim 3[O]$. Subtracting gives

$$[P]+[Q]-2[O] \sim [-R]-[O],$$

which says precisely that the group sum is $-R$. Since the coordinate algorithm preserves k-rational points, the group axioms restrict from $\bar{k}$ to k.

Caution 23.5. A smooth cubic with a chosen point is an elliptic curve, but the simple rule "three collinear points sum to zero" uses a plane embedding in which the chosen identity is a flex. Our Weierstrass embedding has this property. An arbitrary choice of identity on an arbitrary plane model requires an adjusted chord construction.

23.5 A local calculation at infinity

The point O gives a concrete calculation of the orders just defined. In the chart $Y \neq 0$, put $u = X/Y$ and $v = Z/Y$. The equation is

$$v = u^3 + auv^2 + bv^3, \qquad v(1 - auv - bv^2) = u^3.$$

Put $w = 1 - auv - bv^2$. Since $w(O) = 1$, it is a local unit, and

$$v = u^3 w^{-1}, \qquad (u, v) = (u).$$

Thus u is a local parameter. The valuation rules give

$$\begin{aligned} &\operatorname{ord}_O(u) = 1, && \operatorname{ord}_O(v) = 3\operatorname{ord}_O(u) - \operatorname{ord}_O(w) = 3, \\ &\operatorname{ord}_O(x) = \operatorname{ord}_O(u/v) = 1 - 3 = -2, && \operatorname{ord}_O(y) = \operatorname{ord}_O(1/v) = -3. \end{aligned}$$

The double pole of x and triple pole of y explain their respective weights in Weierstrass equations.

Functions with prescribed poles at the identity

The local calculation gives $\operatorname{ord}_O(x) = -2$ and $\operatorname{ord}_O(y) = -3$. Every function regular on $E \setminus \{O\}$ has a unique expression

$$a(x) + yb(x), \qquad a, b \in k[x],$$

because its affine coordinate ring is $k[x,y]/(y^2 - x^3 - ax - b)$, free over $k[x]$ on $1, y$. Nonzero terms from $a(x)$ have even pole orders, while those from $yb(x)$ have odd pole orders at least three. Their leading poles cannot cancel across the two summands. Consequently, for $n \geq 1$, a basis for the space of functions with poles only at O and of order at most n consists of

$$1, \qquad x^i \ (i \geq 1,\ 2i \leq n), \qquad x^j y \ (j \geq 0,\ 2j + 3 \leq n).$$

For example,

$$L(O) = k, \quad L(2O) = \langle 1, x \rangle, \quad L(3O) = \langle 1, x, y \rangle.$$

Counting the displayed monomials gives $\dim L(nO) = n$ for every $n \geq 1$. This is a direct calculation of the special divisors nO in genus-one Riemann–Roch. It does not prove the theorem for arbitrary divisors, which was the additional input used in the divisor-class construction of the group law. The three-dimensional space $L(3O)$ also explains the coordinates $[1 : x : y]$ of a plane cubic model.

23.6 Exercises

Exercise 23.1 (Check the hypotheses before using the formulas)**.** Decide whether $y^2 = x^3 - x$ and $y^2 = x^3 - 3x + 2$ define elliptic curves over $\mathbb{Q}$. Locate any singular point.

Exercise 23.2 (Points of order two)**.** Find all points $P \in E(\mathbb{Q})$ satisfying $2P = O$ for $E : y^2 = x^3 - x$.

Exercise 23.3 (A tangent that meets the same point three times)**.** On $E : y^2 = x^3 + 1$, prove that $P = (0,1)$ has order three.

Exercise 23.4 (Check a doubling computation)**.** On $y^2 = x^3 + 1$, double $Q = (2,3)$ and compare the answer with the earlier sum $P + Q = (-1,0)$, where $P = (0,1)$.

Lecture 24
Rational points, heights, and Mordell's theorem

Rational point computations lead to two questions: whether a point has finite order, and whether finitely many points generate the whole group. Heights measure the size of rational coordinates and provide the setting for descent. References are [2, Chapter IV, §4, final discussion]. The supplementary arithmetic reference is [5, Chapter IV, §§3–4,6]; we prove the descent and height ingredients below, including the arithmetic finiteness lemmas they require.

24.1 Rational point calculations

We now work over $\mathbb{Q}$. For an equation $E : y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Q}$ and nonzero discriminant, $E(\mathbb{Q})$ consists of O and all rational pairs satisfying the equation. Knowing a single point permits us to calculate its multiples, but three distinct tasks must be kept separate:

1. find some rational points;
2. prove that a particular point has infinite order;
3. prove that a proposed list generates every rational point.

The first is a direct calculation. The other two need additional arguments.

Take $E : y^2 = x^3 - 2$ and $P = (3,5)$; indeed $5^2 = 3^3 - 2$. Its tangent slope is $27/10$, and so

$$x(2P) = \frac{729}{100} - 6 = \frac{129}{100}, \qquad y(2P) = \frac{27}{10}\left(3 - \frac{129}{100}\right) - 5 = -\frac{383}{1000}.$$

We have produced a nonintegral rational point. The integral points are generally not closed under the group operation.

To find $3P = P + 2P$, the secant slope is

$$m = \frac{-383/1000 - 5}{129/100 - 3} = \frac{-5383/1000}{-171/100} = \frac{5383}{1710}.$$

Then

$$x(3P) = \left(\frac{5383}{1710}\right)^2 - 3 - \frac{129}{100} = \frac{164323}{29241},$$
$$y(3P) = \frac{5383}{1710}\left(3 - \frac{164323}{29241}\right) - 5 = -\frac{66234835}{5000211}.$$

The fractions grow rapidly. Notice that $29241 = 171^2$ and $5000211 = 171^3$. This denominator pattern has a simple explanation.

Proposition 24.1 (Denominators on an integral Weierstrass model)**.** *Suppose $a, b \in \mathbb{Z}$ and $(x,y) \in E(\mathbb{Q})$. Then*

$$x = \frac{A}{C^2}, \qquad y = \frac{B}{C^3}$$

for some integers A, B, C with $C > 0$, $\gcd(A,C) = \gcd(B,C) = 1$.

Proof. For a prime p, let $v_p(r)$ be the exponent of p in a nonzero rational number r, with numerator exponents counted positively and denominator exponents negatively; set $v_p(0) = +\infty$. For example $v_2(12/5) = 2$ and $v_2(5/12) = -2$. Factoring out the smaller power of p proves that $v_p(r+s)$ equals $\min\{v_p(r), v_p(s)\}$ whenever these two valuations differ. If $v_p(x) = -m < 0$, then

$$v_p(x^3) = -3m < \min\{v_p(ax), v_p(b)\},$$

since $v_p(ax) \geq -m$ and $v_p(b) \geq 0$. The unique smallest valuation cannot cancel, so

$$2v_p(y) = -3m \quad \Longrightarrow \quad m = 2e, \qquad v_p(x) = -2e, \quad v_p(y) = -3e.$$

If $v_p(x) \geq 0$, then $v_p(y) \geq 0$. Set

$$C = \prod_{v_p(x)<0} p^e, \qquad A = C^2 x, \qquad B = C^3 y.$$

These are integers with the asserted coprimalities. Zero coordinates follow from the same integrality argument applied to the curve equation. □

This does not mean homogeneous coordinates are $[A : B : C]$. Clearing denominators in $[x : y : 1]$ gives $[AC : B : C^3]$. A rational projective point can always be represented by a primitive integer triple, meaning a triple of integers whose greatest common divisor is one, unique up to a common sign, but the affine and projective denominator conventions are different.

24.2 Finitely generated groups and Mordell's theorem

An abelian group G, written additively, is *finitely generated* if there are elements $P_1, \dots, P_s$ such that every element is an integer linear combination $\sum_i n_i P_i$. Its *torsion subgroup* is

$$G_{\text{tors}} = \{P \in G : nP = 0 \text{ for some integer } n > 0\}.$$

This is a subgroup: if $mP = 0$ and $nQ = 0$, then $mn(P - Q) = 0$.

The direct sum $\mathbb{Z}^r \oplus T$ consists of pairs with componentwise addition. The integer r is called the *rank* when T is finite. For example $\mathbb{Z}^2 \oplus \mathbb{Z}/6\mathbb{Z}$ has rank two and six torsion elements. The algebraic structure theorem needed for Mordell's theorem is the following.

Proposition 24.2 (Structure of finitely generated abelian groups)**.** *Every finitely generated abelian group is isomorphic to $\mathbb{Z}^r \oplus T$ for a uniquely determined integer $r \geq 0$ and a finite abelian group T. Under this isomorphism T is its torsion subgroup.*

Proof. Choose d generators of G and map $\mathbb{Z}^d$ onto G by sending the standard basis to them. Its kernel N is finitely generated because $\mathbb{Z}$ is Noetherian and N is a submodule of the finite free module $\mathbb{Z}^d$. A choice of generators of N gives an integer matrix whose columns are those generators. Changing a basis of $\mathbb{Z}^d$ amounts to invertible integer row operations. Invertible column operations change the chosen generators of N without changing its image.

The Euclidean algorithm reduces this matrix to a diagonal matrix with nonzero diagonal entries $a_1, \dots, a_t$. Here are the steps that ensure the reduction terminates. Move a nonzero entry to the upper left and make it positive. Division with remainder, using row operations, reduces the other entries in its column; column operations do the same in its row. Whenever a nonzero remainder occurs, move it to the upper left. The positive pivot strictly decreases, so eventually it divides every entry in its row and column. Subtract multiples to clear that row and column. If the pivot does not divide an entry of the remaining block, add that entry's row to the first row. A further column division now produces a smaller positive pivot, so restart. Eventually the pivot divides the entire remaining block. Apply the procedure to that block, leaving the cleared row and column fixed. Induction on the matrix size gives the claimed diagonal form. This procedure is the *Smith normal form* algorithm.

For the diagonal presentation, the quotient is

$$G \cong \mathbb{Z}^{d-t} \oplus \bigoplus_{i=1}^{t} \mathbb{Z}/a_i\mathbb{Z}.$$

The second summand is finite, and it is exactly the torsion subgroup because a nonzero integer vector has infinite order. Finally $\mathbb{Q} \otimes_{\mathbb{Z}} G$ is a vector space of dimension $d - t$: tensoring kills each finite cyclic summand and sends $\mathbb{Z}$ to $\mathbb{Q}$. Hence $r = d - t$ depends only on G. □

The assertion specific to elliptic curves is their finite generation.

Theorem 24.3 (Mordell)**.** *For an elliptic curve E over $\mathbb{Q}$, the abelian group $E(\mathbb{Q})$ is finitely generated. Consequently*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \oplus T$$

for a nonnegative integer r and a finite abelian group T.

The displayed decomposition follows by applying the structure theorem to $E(\mathbb{Q})$. A *number field* is a finite field extension of $\mathbb{Q}$, such as $\mathbb{Q}(\sqrt{2})$. The analogous finite-generation assertion over any number field is called the Mordell–Weil theorem.

We prove this theorem in the remainder of the lecture. We use the elliptic-curve group law from Lecture 23 and the ideal theory of Lecture 10. The new arithmetic ingredients are proved here. The argument has two parts: first, there are only finitely many rational points up to adding doubles; second, a point of large arithmetic size can be replaced by a smaller point in a controlled way. The intermediate number field is an auxiliary device; the theorem we prove concerns all elliptic curves over $\mathbb{Q}$, whether or not their points of order two are rational.

We may use an integral equation. Choose a positive integer d divisible by the denominators of a, b. Then

$$\begin{aligned} x' &= d^2x, & y' &= d^3y, \\ y'^2 &= x'^3 + a'x' + b', & a' &= d^4a \in \mathbb{Z}, \quad b' = d^6b \in \mathbb{Z}, \\ \Delta' &= d^{12}\Delta \neq 0. \end{aligned}$$

The coordinate change and its inverse preserve lines, tangency, reflection, and O, hence induce an isomorphism on groups of rational points. Rename a', b' as a, b.

Caution 24.4. A computation of $P, 2P, 3P$ does not determine the rank. Even proving P has infinite order would give only $r \geq 1$. To prove $r = 1$, one must also rule out further independent points. We make no rank or generator claim for $y^2 = x^3 - 2$ on the basis of the calculations above.

The logical structure of the finite-generation argument

The proof below has three distinct finiteness steps. First one proves that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. This gives finitely many representatives R_i, so every point can be written $P = R_i + 2Q$. It does not yet bound Q or produce a finite generating set. For example the additive group $\mathbb{Q}$ satisfies $\mathbb{Q}/2\mathbb{Q} = 0$ but is not finitely generated: finitely many rationals have a common denominator, while the group contains rationals with arbitrarily large new denominators.

Second, a height measures the arithmetic size of a point and has only finitely many points below any fixed bound. Third, estimates for doubling and translation turn $P = R_i + 2Q$ into

$$h_E(Q) \leq \tfrac{1}{2} h_E(P) + C.$$

Only this final inequality converts the finite quotient into finite generation. It forces large points to be expressed using smaller ones, until a bounded set of points suffices.

The arithmetic part of weak Mordell has its own division of labor. The splitting field makes the three roots e_i available; the square-class maps detect divisibility by two; valuations show that only finitely many prime ideals can occur with odd exponent; finiteness of ideal classes and of units modulo squares then bounds the possible square classes. Keeping

these steps separate prevents the incorrect inference that bounded denominators alone prove finite generation.

24.3 Heights and descent

For a reduced rational number $t = u/v$ with $v > 0$, define

$$H(t) = \max\{|u|, v\}, \qquad h(t) = \log H(t).$$

This is not the absolute value $|t|$. For example $1000/1001$ is close to 1 as a real number, but has height $\log 1001$, whereas 1 has height zero. A height measures arithmetic complexity, not distance from zero.

Set $h_E(P) = h(x(P))$ for affine P and $h_E(O) = 0$. For $B \geq 0$, put $N = \lfloor e^B \rfloor$. Then

$$h_E(P) \leq B, \quad x(P) = u/v \text{ reduced} \quad \Longrightarrow \quad |u| \leq N, \quad 1 \leq v \leq N.$$

Each x permits at most two values of y, so

$$\#\{P \in E(\mathbb{Q}) : h_E(P) \leq B\} \leq 1 + 2N(2N+1) < \infty.$$

For $B < 0$ the set is empty.

24.4 The arithmetic needed for descent

The cubic defining an elliptic curve need not have its roots in $\mathbb{Q}$. We will therefore work for a while in a number field K, a finite extension of $\mathbb{Q}$, and then return to rational points. The arithmetic question is elementary to state: if the prime exponents of a nonzero element are even except at finitely many specified primes, how many possibilities remain after multiplication by squares? We will prove that only finitely many remain. All the finiteness arguments are supplied below.

Write $n = [K : \mathbb{Q}]$ and let $\mathcal{O}_K$ be the ring of algebraic integers in K. By Proposition 10.29 and Theorem 10.44, this is a finite $\mathbb{Z}$-module, it spans K over $\mathbb{Q}$, and it is a Dedekind domain. It has no additive torsion, because it is contained in a field of characteristic zero. The structure theorem for finitely generated abelian groups therefore gives a $\mathbb{Z}$-basis

$$\omega_1, \ldots, \omega_n \quad \text{of } \mathcal{O}_K.$$

Such a basis is called an *integral basis*. The number of basis elements is n because their $\mathbb{Q}$-span is K. In particular, each algebraic integer has a unique expression $\sum_i c_i \omega_i$ with $c_i \in \mathbb{Z}$.

Recall from Lecture 10 that each nonzero fractional ideal has a unique factorization into powers of nonzero prime ideals. For $a \in K^\times$, write

$$(a) = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(a)}.$$

Here the integers $v_{\mathfrak{p}}(a)$ may be negative, and only finitely many are nonzero. They are the valuations of the DVRs $(\mathscr{O}_K)_{\mathfrak{p}}$, so they satisfy the valuation rules proved in Lecture 10. In particular, $a \in \mathscr{O}_K$ exactly when every exponent is nonnegative, and a is a unit exactly when every exponent is zero. For example, in $\mathbb{Q}$ this is the usual prime factorization of a nonzero rational number.

24.4.1 Norms and a pigeonhole argument

There are exactly n embeddings $\sigma_1, \ldots, \sigma_n : K \to \mathbb{C}$ fixing $\mathbb{Q}$, by Proposition 10.28. The *field norm* of $a \in K$ is

$$\mathrm{N}_{K/\mathbb{Q}}(a) = \det(m_a), \qquad m_a : K \longrightarrow K, \quad x \longmapsto ax.$$

The determinant is taken over $\mathbb{Q}$. The diagonalization used to prove the trace formula in Lecture 10 also proves

$$\mathrm{N}_{K/\mathbb{Q}}(a) = \prod_{i=1}^{n} \sigma_i(a).$$

The values $\sigma_i(a)$ are the *conjugates* of a, here counted with the multiplicities arising from the n embeddings. For an algebraic integer a, multiplication by a has an integer matrix in an integral basis. Its characteristic polynomial consequently lies in $\mathbb{Z}[T]$, and its determinant is an integer.

For a nonzero ordinary ideal $I \subseteq \mathscr{O}_K$, its *ideal norm* is the number of elements of its quotient:

$$\mathrm{N}(I) = |\mathscr{O}_K/I|.$$

This number is finite. Indeed, choose $0 \neq a \in I$. Multiplication by a is an injective map of the free abelian group $\mathscr{O}_K$ to itself, with nonzero integer determinant. Smith normal form shows that its cokernel has order the absolute value of that determinant. Thus

$$\mathrm{N}((a)) = |\mathrm{N}_{K/\mathbb{Q}}(a)|, \tag{24.1}$$

and $\mathscr{O}_K/I$, being a quotient of $\mathscr{O}_K/(a)$, is finite as well. The two norms have different definitions; equation (24.1) explains their relationship.

Lemma 24.5 (Multiplication of ideal norms)**.** *For nonzero ordinary ideals I, J of $\mathscr{O}_K$,*

$$\mathrm{N}(IJ) = \mathrm{N}(I)\mathrm{N}(J).$$

Proof. We first consider a nonzero prime $\mathfrak{p}$. The quotient $k_{\mathfrak{p}} = \mathscr{O}_K/\mathfrak{p}$ is a finite field. Choose $\pi \in \mathscr{O}_K$ with $v_{\mathfrak{p}}(\pi) = 1$: a uniformizer in $(\mathscr{O}_K)_{\mathfrak{p}}$ is a fraction r/s with $s \notin \mathfrak{p}$, and its numerator r has this property. For every $j \geq 0$, multiplication by π^j induces an isomorphism

$$k_{\mathfrak{p}} \longrightarrow \mathfrak{p}^{j}/\mathfrak{p}^{j+1}.$$

Here is why a local generator suffices for this assertion. The target is a vector space over $k_{\mathfrak{p}}$, so multiplication by each $s \notin \mathfrak{p}$ is already invertible on it; localizing at $\mathfrak{p}$ does not change it. In the DVR $(\mathcal{O}_K)_{\mathfrak{p}}$ the displayed map is the usual identification of $(\pi^j)/(\pi^{j+1})$ with the residue field, so it is an isomorphism before localization too. Consequently the filtration of $\mathcal{O}_K/\mathfrak{p}^e$ by consecutive powers of $\mathfrak{p}$ has e quotients of size $\mathrm{N}(\mathfrak{p})$, giving

$$\mathrm{N}(\mathfrak{p}^e) = \mathrm{N}(\mathfrak{p})^e.$$

Distinct nonzero primes are distinct maximal ideals. Their powers are comaximal, so the Chinese remainder theorem and ideal factorization give

$$\mathrm{N}\left(\prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}}\right) = \prod_{\mathfrak{p}} \mathrm{N}(\mathfrak{p})^{e_{\mathfrak{p}}}.$$

Multiplication of ideals adds the exponents. The desired formula follows. □

Lemma 24.6 (A small element in an ideal)**.** *There is a constant $C_K \geq 1$, depending only on K and a chosen integral basis, such that every nonzero ordinary ideal I contains an element $a \neq 0$ satisfying*

$$|\mathrm{N}_{K/\mathbb{Q}}(a)| \leq C_K \mathrm{N}(I).$$

Proof. Put $m = \mathrm{N}(I)$ and $T = \lfloor m^{1/n} \rfloor$, where $\lfloor t \rfloor$ denotes the greatest integer not exceeding t. Consider the $(T+1)^n$ algebraic integers

$$\sum_{i=1}^{n} b_i \omega_i, \qquad 0 \leq b_i \leq T, \quad b_i \in \mathbb{Z}.$$

They are distinct because the ω_i are a basis. Since $(T+1)^n > m$, two have the same image in the m-element ring $\mathcal{O}_K/I$. Their difference is a nonzero element

$$a = \sum_{i} c_i \omega_i \in I, \qquad |c_i| \leq T.$$

For each embedding, the triangle inequality gives

$$|\sigma_j(a)| \leq T \sum_{i} |\sigma_j(\omega_i)|.$$

Multiply these n inequalities. Since $T^n \leq m$, the norm formula gives the required bound with

$$C_K = \max\left\{1, \prod_{j=1}^{n} \left(\sum_{i=1}^{n} |\sigma_j(\omega_i)| \right) \right\}.$$

The constant does not depend on the ideal. The argument uses only the pigeonhole principle on an integer box. □

Proposition 24.7 (Finiteness of the ideal class group)**.** *The ideal class group* $\mathrm{Cl}(\mathscr{O}_K)$ *is finite.*

Proof. Recall that two nonzero fractional ideals have the same class if one is a nonzero scalar multiple of the other. Products of ideals give the group operation, and a principal ideal has the identity class. Every class has an ordinary ideal representative, obtained by clearing denominators.

Let c be any class, and choose an ordinary ideal I representing c^{-1}. The preceding lemma supplies $0 \neq a \in I$ with $|\mathrm{N}_{K/\mathbb{Q}}(a)| \leq C_K \mathrm{N}(I)$. Because $a \in I$ and $II^{-1} = \mathscr{O}_K$, the ideal

$$J = (a)I^{-1}$$

is an ordinary ideal: it is contained in $II^{-1} = \mathscr{O}_K$. It represents the class c. Moreover $IJ = (a)$, and the two norm formulas give

$$\mathrm{N}(J) = \frac{|\mathrm{N}_{K/\mathbb{Q}}(a)|}{\mathrm{N}(I)} \leq C_K.$$

Thus every class has an ordinary ideal representative of bounded norm.

There are only finitely many such ideals. To see this directly, put $B = \lceil C_K \rceil$. If $\mathrm{N}(J) = m \leq B$, Lagrange's theorem gives

$$m(\mathscr{O}_K/J) = 0 \quad \Longrightarrow \quad B!\,\mathscr{O}_K \subseteq m\mathscr{O}_K \subseteq J.$$

Thus all such J are inverse images of ideals of the finite ring $\mathscr{O}_K/(B!\,\mathscr{O}_K)$. A finite ring has only finitely many subsets, hence only finitely many ideals. This proves the proposition. □

24.4.2 Units modulo squares

Write $U = \mathscr{O}_K^\times$ for the multiplicative group of units. For any abelian multiplicative group G, its squares form the subgroup $G^2 = \{g^2 : g \in G\}$. The quotient G/G^2 records elements up to multiplication by a square. We next prove that U/U^2 is finite. We do not need a formula for the number of independent units.

Lemma 24.8 (Bounded conjugates)**.** *For every real number* $B > 0$*, there are only finitely many* $a \in \mathscr{O}_K$ *such that* $|\sigma_i(a)| \leq B$ *for all* i*.*

Proof. The characteristic polynomial of multiplication by a is

$$F_a(T) = \prod_{i=1}^{n} (T - \sigma_i(a)) \in \mathbb{Z}[T].$$

Writing $F_a(T) = T^n + c_1 T^{n-1} + \cdots + c_n$, its coefficients satisfy

$$c_j = (-1)^j \sum_{i_1 < \cdots < i_j} \sigma_{i_1}(a) \cdots \sigma_{i_j}(a) \in \mathbb{Z}, \qquad |c_j| \leq \binom{n}{j} B^j.$$

Each c_j ranges over finitely many integers, so only finitely many F_a are possible. Fix one embedding $K \hookrightarrow \mathbb{C}$. The image of a under this embedding is a root of F_a, and each of these finitely many nonzero degree-n polynomials has at most n roots. Injectivity of the embedding proves finiteness of the possible a. □

Lemma 24.9 (A bounded-set criterion for finite generation)**.** *Let Γ be an additive subgroup of $\mathbb{R}^n$. Suppose that its intersection with every bounded subset of $\mathbb{R}^n$ is finite. Then Γ is finitely generated as an abelian group.*

Proof. Choose $\gamma_1, \dots, \gamma_r \in \Gamma$ forming a real basis of the real span of Γ; this is possible by finite-dimensional linear algebra. If the span is zero, then $\Gamma = 0$ and there is nothing to prove. Otherwise consider the bounded parallelepiped

$$D = \left\{ \sum_{i=1}^{r} t_i \gamma_i : 0 \le t_i \le 1 \right\}.$$

Its intersection with Γ is finite. For $\gamma = \sum_i a_i \gamma_i \in \Gamma$,

$$\gamma = \sum_i \lfloor a_i \rfloor \gamma_i + \eta, \qquad \eta = \sum_i (a_i - \lfloor a_i \rfloor)\gamma_i \in \Gamma \cap D.$$

Therefore

$$\Gamma = \langle \gamma_1, \dots, \gamma_r, \Gamma \cap D \rangle_{\mathbb{Z}},$$

a group generated by finitely many elements. □

Proposition 24.10 (Finiteness of unit square classes)**.** *The group $U = \mathscr{O}_K^\times$ is finitely generated. In particular, U/U^2 is finite.*

Proof. Use the ordinary real logarithm to define

$$L : U \longrightarrow \mathbb{R}^n, \qquad L(u) = \big(\log|\sigma_1(u)|, \dots, \log|\sigma_n(u)|\big).$$

No logarithm of zero occurs because a unit is nonzero and each embedding is injective. Since $\log|zw| = \log|z| + \log|w|$, this is a homomorphism from a multiplicative group to an additive group.

If the coordinates of $L(u)$ are bounded in absolute value by M, then all conjugates of u have absolute value at most e^M. Lemma 24.8 gives only finitely many such units. Consequently $L(U)$ meets every bounded set in a finite set, and it is finitely generated by Lemma 24.9. The same bounded-conjugate argument, with $B = 1$, shows that $\ker L$ is finite.

Choose units lifting a finite generating set of $L(U)$. For any unit u, its logarithm is an integer combination of the chosen logarithms. Dividing u by the corresponding product of the chosen units leaves an element of $\ker L$. These finitely many lifts, together with the finite kernel, generate U. Finally, if $u_1, \dots, u_s$ generate U, every unit modulo U^2 is represented by

$$u_1^{\varepsilon_1} \cdots u_s^{\varepsilon_s}, \qquad \varepsilon_i \in \{0, 1\}.$$

There are at most 2^s such representatives. □

24.4.3 Finitely many square classes with restricted prime exponents

Two elements $a, b \in K^\times$ have the same *square class* if $a/b = t^2$ for some $t \in K^\times$. Denote the class of a by $[a]$ in $K^\times/(K^\times)^2$. If S is a finite set of nonzero prime ideals of $\mathcal{O}_K$, define

$$K(S,2) = \left\{ [a] \in K^\times/(K^\times)^2 : v_\mathfrak{p}(a) \text{ is even whenever } \mathfrak{p} \notin S \right\}.$$

This condition is independent of the representative because multiplication by t^2 adds $2v_\mathfrak{p}(t)$ to every exponent. The set is a subgroup, since the valuation of a product is the sum of the valuations.

For $K = \mathbb{Q}$ and S consisting of the primes 2 and 3, representatives are

$$1, -1, 2, -2, 3, -3, 6, -6.$$

Every other prime occurs to an even power and can be removed by a rational square; powers of 2 and 3 can be reduced to exponents 0 or 1. In a general number field, ideals need not be principal, and units need not be just $1, -1$. The two finiteness propositions above account for exactly these complications.

Proposition 24.11 (Finiteness of restricted square classes)**.** *For every number field K and every finite set S of nonzero primes of $\mathcal{O}_K$, the group $K(S,2)$ is finite.*

Proof. Choose a representing a class in $K(S,2)$. Reduce the exponents at primes in S modulo 2, writing their remainders as $\varepsilon_\mathfrak{p} \in \{0,1\}$. Unique factorization of fractional ideals gives

$$(a) = \mathfrak{b}\,\mathfrak{c}^2, \qquad \mathfrak{b} = \prod_{\mathfrak{p} \in S} \mathfrak{p}^{\varepsilon_\mathfrak{p}},$$

where $\mathfrak{c}$ is a nonzero fractional ideal. All exponents in $\mathfrak{c}$ are integers because we removed the parity at the primes in S, and the other exponents were even already.

There are at most $2^{|S|}$ possibilities for $\mathfrak{b}$, and only finitely many possibilities for the ideal class of $\mathfrak{c}$, by Proposition 24.7. For each pair $(\mathfrak{b}, [\mathfrak{c}])$ that occurs, choose one element a_0 with

$$(a_0) = \mathfrak{b}\,\mathfrak{c}_0^2.$$

If a gives the same pair, then $\mathfrak{c}$ and $\mathfrak{c}_0$ have the same class. Thus $\mathfrak{c} = (t)\mathfrak{c}_0$ for some $t \in K^\times$. It follows that

$$(a) = (t^2 a_0), \qquad a = t^2 a_0 u$$

for a unit $u \in U$: equality of the principal ideals says that the quotient and its inverse are both in $\mathcal{O}_K$. Consequently $[a] = [a_0 u]$. Proposition 24.10 supplies only finitely many possibilities for u up to a square of a unit, hence up to a square in K. Finitely many chosen pairs and finitely many unit square classes give finitely many classes $[a]$, as required. □

We have proved the entire arithmetic finiteness statement needed below. Its ingredients were ideal factorization from Lecture 10, integer matrices, the pigeonhole principle, and bounded sets in a finite-dimensional real vector space.

24.5 Weak Mordell: reducing to finitely many square classes

For an abelian group G, let $2G = \{2P : P \in G\}$. It is a subgroup, and two elements have the same image in $G/2G$ precisely when their difference is twice an element of G. We shall prove that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. This assertion is called the *weak Mordell theorem*. Its role is to supply finitely many starting points for the height descent later in the lecture.

We continue with the integral equation $E : y^2 = f(x) = x^3 + ax + b$ fixed above.

24.5.1 A field in which the three roots are visible

Choose the three distinct complex roots e_1, e_2, e_3 of f, and put

$$K = \mathbb{Q}(e_1, e_2, e_3).$$

This is the *splitting field*: the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and all the roots. It is a number field. In fact adjoining e_1 has degree at most three, adjoining a second root then has degree at most two, and $e_3 = -e_1 - e_2$; thus $[K : \mathbb{Q}] \leq 6$. Each e_i is an algebraic integer because it satisfies the monic polynomial $f \in \mathbb{Z}[X]$.

A $\mathbb{Q}$-*automorphism* of K is a field isomorphism $\sigma : K \to K$ that fixes every rational number. These automorphisms form a group Γ under composition. Each permutes the three roots, and its values on those roots determine it, so Γ is finite, with at most six elements. We need one further property:

Lemma 24.12 (The elements fixed by all automorphisms)**.** *If $z \in K$ satisfies $\sigma(z) = z$ for every $\sigma \in \Gamma$, then $z \in \mathbb{Q}$.*

Proof. Suppose $z \notin \mathbb{Q}$. Its minimal polynomial over $\mathbb{Q}$ is irreducible and has degree greater than one. In characteristic zero it has no repeated root: its derivative is nonzero and has smaller degree, so irreducibility implies that the polynomial and its derivative are relatively prime. Choose another complex root z'. Evaluation gives an isomorphism $\mathbb{Q}(z) \to \mathbb{Q}(z')$ carrying z to z'.

We explain why this isomorphism extends to an embedding of K into $\mathbb{C}$. If an embedding of a field F into $\mathbb{C}$ is already given and θ is algebraic over F, apply the embedding to the coefficients of the minimal polynomial of θ. Choose a complex root of the resulting polynomial. Sending θ to that root extends the embedding: the quotient description $F[\theta] \cong F[T]/(m_\theta)$ proves that this map is well defined and injective. Apply this construction successively to the finitely many generators e_1, e_2, e_3 of K over $\mathbb{Q}(z)$.

The resulting embedding sends the three distinct roots of f to three distinct roots of the same polynomial. It therefore permutes them, and its image is exactly K. It is an element of Γ taking z to $z' \neq z$, a contradiction. □

The term *finite Galois extension* applies to this splitting field in characteristic zero; the lemma states the fixed-field property we shall use. No other Galois-theoretic result is needed below. An automorphism acts on points by

$$\sigma(x,y) = (\sigma(x), \sigma(y)), \qquad \sigma(O) = O.$$

The addition formulas have rational coefficients, so $\sigma(P+Q) = \sigma(P) + \sigma(Q)$. By the lemma, a point of $E(K)$ fixed by every σ belongs to $E(\mathbb{Q})$.

24.5.2 The square-class map

Recall that the *square class* $[u]$ of $u \in K^\times$ is its coset in $K^\times/K^{\times 2}$: two nonzero elements have the same class when their quotient is a square. Multiplication of representatives defines the group operation, and every class has square $[1]$.

Over K our equation is

$$y^2 = (x-e_1)(x-e_2)(x-e_3).$$

Its points killed by 2 are exactly

$$E(K)[2] := \{P : 2P = O\} = \{O, T_1, T_2, T_3\}, \qquad T_i = (e_i, 0).$$

To check this, recall that $-P = (x,-y)$. An affine point satisfies $P = -P$ exactly when $y = 0$, because the characteristic is zero. These are precisely the three displayed points.

For $\{i,j,k\} = \{1,2,3\}$ define

$$\delta_i(P) = \begin{cases} [1], & P = O, \\ [(e_i - e_j)(e_i - e_k)], & P = T_i, \\ [x(P) - e_i], & P \neq O, T_i. \end{cases}$$

The special value at T_i avoids taking the square class of zero. The following calculation explains why this particular replacement is the correct one.

Lemma 24.13 (The chord calculation)**.** *Each map $\delta_i : E(K) \to K^\times/K^{\times 2}$ is a group homomorphism.*

Proof. Fix i. We first show that if a line meets the cubic in P, Q, R, with intersections repeated in the tangent case, then

$$\delta_i(P)\delta_i(Q)\delta_i(R) = [1].$$

For a nonvertical affine line $y = mX + c$, substitution gives the monic cubic identity

$$f(X) - (mX+c)^2 = (X - x(P))(X - x(Q))(X - x(R)).$$

If none of the three points is T_i, evaluating at $X = e_i$ gives

$$(x(P) - e_i)(x(Q) - e_i)(x(R) - e_i) = (me_i + c)^2.$$

All factors are nonzero, so their product represents the trivial square class. This identity also applies when a root is repeated.

If the line contains T_i, rename that point R. Then $me_i + c = 0$. The intersection at T_i is simple: differentiating the left side at e_i gives $f'(e_i) \neq 0$. Differentiating the right side now gives

$$(e_i - x(P))(e_i - x(Q)) = f'(e_i) = (e_i - e_j)(e_i - e_k).$$

Multiplication by the special value used in $\delta_i(T_i)$ produces the square $f'(e_i)^2$. This covers even the line $y = 0$ through all three points of order two.

A vertical line meets the cubic in $P, -P, O$, with P repeated when $P = -P$. Since $\delta_i(-P) = \delta_i(P)$ and every square class has square $[1]$, the required identity holds here as well. The line at infinity has three intersections at O and gives the same conclusion.

For the group law, the third point on the chord or tangent through P, Q is $R = -(P + Q)$. Since $\delta_i(-R) = \delta_i(R)$, the identity proves $\delta_i(P + Q) = \delta_i(P)\delta_i(Q)$. If either point is O, this equality follows directly from the definition. □

Combine the three maps into

$$\delta = (\delta_1, \delta_2, \delta_3) : E(K) \longrightarrow (K^\times / K^{\times 2})^3.$$

The next step identifies its kernel. We give an explicit construction, so that the claim that a point can be halved does not hide an existence theorem.

Lemma 24.14 (An explicit half of a point)**.** *The kernel of δ is $2E(K)$.*

Proof. Since δ is a homomorphism and the square of every element of its target is the identity, $\delta(2Q) = ([1], [1], [1])$ for every Q. Thus $2E(K)$ is contained in the kernel.

Conversely, suppose $P \in \ker \delta$. If $P = O$, it is $2O$. Suppose $P = (x, y)$ is affine. We can choose $r_1, r_2, r_3 \in K$ such that

$$r_i^2 = x - e_i, \qquad r_1 r_2 r_3 = -y.$$

Indeed, if P is not a point of order two, the three square roots exist by the definition of δ. Their product has square y^2, so changing one sign makes the product $-y$. If $P = T_i$, take $r_i = 0$; the other two square roots exist because $\delta_j(T_i)$ and $\delta_k(T_i)$ are trivial, and the product condition is automatic.

Put

$$s_1 = r_1 + r_2 + r_3, \qquad s_2 = r_1 r_2 + r_1 r_3 + r_2 r_3,$$

and define

$$Q = (x + s_2,\ -y - s_1 s_2).$$

We check both that Q is on the curve and that $2Q = P$. Write

$$D = (r_1 + r_2)(r_1 + r_3)(r_2 + r_3).$$

Expanding the product gives $D = s_1 s_2 - r_1 r_2 r_3 = s_1 s_2 + y$, so $y(Q) = -D$. Also

$$x(Q) - e_i = r_i^2 + s_2 = (r_i + r_j)(r_i + r_k).$$

Multiplying these three identities gives $f(x(Q)) = D^2 = y(Q)^2$. No factor $r_i + r_j$ is zero, because that would imply $r_i^2 = r_j^2$, and hence $e_i = e_j$. Thus $D \neq 0$, so $y(Q) \neq 0$ and the ordinary tangent formula applies.

Differentiate $f(X) = \prod_i (X - e_i)$ and evaluate at $x(Q)$. Each term of $f'(x(Q))$ is D times one of $r_1 + r_2$, $r_1 + r_3$, $r_2 + r_3$. Consequently

$$f'(x(Q)) = 2s_1 D, \qquad m = \frac{f'(x(Q))}{2y(Q)} = -s_1.$$

Finally, $e_1 + e_2 + e_3 = 0$ implies

$$s_1^2 = r_1^2 + r_2^2 + r_3^2 + 2s_2 = 3x + 2s_2.$$

The doubling formulas therefore give

$$\begin{aligned} x(2Q) &= m^2 - 2x(Q) = s_1^2 - 2(x + s_2) = x, \\ y(2Q) &= m(x(Q) - x) - y(Q) = -s_1 s_2 + y + s_1 s_2 = y. \end{aligned}$$

Hence $2Q = P$, including when P has order two. This proves the reverse inclusion. □

24.5.3 Only finitely many square classes occur

Let $\mathcal{O}_K$ be the ring of integers of K, and let S be the finite set of its nonzero prime ideals dividing the nonzero principal ideal

$$\left(2 \prod_{i<j} (e_i - e_j) \right).$$

Here a prime ideal *divides* an ideal when it occurs in its prime ideal factorization. For $\mathfrak{p} \notin S$, every difference $e_i - e_j$ has valuation zero. We use the notation $v_\mathfrak{p}$ and the finite group $K(S,2)$ established in Proposition 24.11.

Lemma 24.15 (Valuations away from a finite set)**.** *For every $P \in E(K)$ and every i, the class $\delta_i(P)$ belongs to $K(S,2)$.*

Proof. Fix $\mathfrak{p} \notin S$. At O the representative is 1. At T_j each representative in the definition of δ_i is a root difference or a product of two root differences, so its valuation is zero.

Now let $P = (x, y)$ be another affine point. If $v_\mathfrak{p}(x) < 0$, integrality of the e_i gives

$$v_\mathfrak{p}(x - e_i) = v_\mathfrak{p}(x) \quad \text{for all } i.$$

The equation then implies $2v_\mathfrak{p}(y) = 3v_\mathfrak{p}(x)$, so $v_\mathfrak{p}(x)$ is even. If $v_\mathfrak{p}(x) \geq 0$, all three $x - e_i$ have nonnegative valuation. At most one can have positive valuation: two such factors would have a difference $e_j - e_i$ of positive valuation, contrary to $\mathfrak{p} \notin S$. Their product is y^2, so the sum of their valuations is even. The sole possibly positive valuation is therefore even

too. Thus every representative has even valuation outside S, exactly the defining condition for $K(S,2)$. □

By Lemma 24.14 and the first isomorphism theorem, δ induces an injection

$$E(K)/2E(K) \hookrightarrow K(S,2)^3.$$

The group on the right is finite by Proposition 24.11. Hence $E(K)/2E(K)$ is finite.

24.5.4 Returning from K to $\mathbb{Q}$

A rational point that is twice a K-point need not be twice a rational point. Thus the inclusion $E(\mathbb{Q}) \subseteq E(K)$ does not automatically give an injection of the quotients by 2. The following finite counting argument supplies the required missing step.

Example 24.16 (Halving can require a larger field)**.** On $E : y^2 = x^3 - x$, let $T = (1,0)$ and order the roots as $e_1 = -1$, $e_2 = 0$, $e_3 = 1$. Then

$$\delta(T) = ([2],[1],[2]).$$

Since 2 is not a square in $\mathbb{Q}$, Lemma 24.14 shows that $T \notin 2E(\mathbb{Q})$. Over $L = \mathbb{Q}(\sqrt{2})$, however, take $r_1 = \sqrt{2}$, $r_2 = 1$, $r_3 = 0$. The same lemma gives

$$Q = (1+\sqrt{2}, -2-\sqrt{2}), \qquad 2Q = T.$$

Thus a nonzero class in $E(\mathbb{Q})/2E(\mathbb{Q})$ becomes zero in $E(L)/2E(L)$. This example explains why we must prove finiteness of the kernel when passing between fields.

Theorem 24.17 (Weak Mordell over $\mathbb{Q}$)**.** *For every elliptic curve E over $\mathbb{Q}$, the quotient $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.*

Proof. Consider the homomorphism

$$\rho : E(\mathbb{Q})/2E(\mathbb{Q}) \longrightarrow E(K)/2E(K)$$

induced by inclusion. Its image is finite by the preceding argument. We prove that its kernel is finite too.

An element of the kernel has a representative $P \in E(\mathbb{Q})$ with $P = 2Q$ for some $Q \in E(K)$. For every $\sigma \in \Gamma$, the difference

$$c_Q(\sigma) = \sigma(Q) - Q$$

is killed by two, since $2c_Q(\sigma) = \sigma(P) - P = O$. Thus c_Q is a function from the finite set Γ to the four-element set $E(K)[2]$. There are at most $4^{|\Gamma|}$ such functions.

Suppose two choices $P = 2Q$ and $P' = 2Q'$ give the same function. Then, for every σ,

$$\sigma(Q - Q') = (Q + c_Q(\sigma)) - (Q' + c_{Q'}(\sigma)) = Q - Q'.$$

The fixed-field lemma implies $Q - Q' \in E(\mathbb{Q})$. Hence $P - P' = 2(Q - Q') \in 2E(\mathbb{Q})$, so P and P' represent the same element of the kernel. Therefore a given function can arise from at most one kernel element. Although a kernel element may admit several choices of half, this observation still bounds the number of kernel elements by $4^{|\Gamma|}$.

Both the kernel and image of ρ are finite. Each nonempty fiber is a coset of its kernel and therefore has the same finite size. There are only finitely many such fibers, proving that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. □

The square-class calculation is a form of *descent by* 2. It replaces the question of dividing points by two with arithmetic in a finite set of square classes. Milne discusses this approach in [5, Chapter IV, Remark 3.7]; the calculations above include the homomorphism, halving, and field-extension arguments explicitly.

24.6 Height estimates and the proof of Mordell's theorem

24.6.1 How polynomial formulas change height

We now prove the estimates that make descent possible. We continue to use the integral Weierstrass equation fixed above.

It is convenient to extend the height to the projective line. Write a rational projective point as $[m : n]$ with $m, n \in \mathbb{Z}$, not both zero, and $\gcd(m, n) = 1$. Such a pair is unique up to simultaneous change of sign. Define

$$H([m : n]) = \max\{|m|, |n|\}, \qquad h([m : n]) = \log H([m : n]).$$

Thus $h([t : 1]) = h(t)$ for $t \in \mathbb{Q}$, and $h([1 : 0]) = 0$. If we set $x(O) = [1 : 0]$, then $h_E(P) = h(x(P))$ for every point, including O.

Lemma 24.18 (Height of a polynomial map)**.** *Let $F, G \in \mathbb{Z}[X, Z]$ be homogeneous polynomials of the same positive degree d. Suppose they have no common zero in $\mathbb{P}^1(\mathbb{C})$. Then the map*

$$\varphi : \mathbb{P}^1(\mathbb{Q}) \longrightarrow \mathbb{P}^1(\mathbb{Q}), \qquad [m : n] \longmapsto [F(m, n) : G(m, n)]$$

is well-defined, and there is a constant $c \geq 0$, depending only on F, G, such that

$$\left| h(\varphi(T)) - d\, h(T) \right| \leq c \qquad (T \in \mathbb{P}^1(\mathbb{Q})).$$

Proof. Homogeneity makes the output independent of a simultaneous rescaling of m, n. The hypothesis ensures that its two coordinates are not both zero. There are two issues in estimating its height: the size of the polynomial values, and the common factor cancelled when we reduce the output coordinates.

Step 1: cancellation is uniformly bounded. The polynomials $F(T, 1)$ and $G(T, 1)$ have no common complex root, so they are relatively prime in $\mathbb{Q}[T]$. The Euclidean algorithm gives polynomials $U, V \in \mathbb{Q}[T]$ with

$$U(T)F(T,1)+V(T)G(T,1)=1.$$

Clear their coefficient denominators and then substitute $T=X/Z$. Multiplying by a sufficiently large power of Z clears all negative exponents. We obtain an identity

$$U_1(X,Z)F(X,Z)+V_1(X,Z)G(X,Z)=D_1Z^{N_1},$$

where $U_1,V_1\in\mathbb{Z}[X,Z]$, D_1 is a nonzero integer, and $N_1>0$. Applying the same argument to $F(1,T),G(1,T)$ gives

$$U_2(X,Z)F(X,Z)+V_2(X,Z)G(X,Z)=D_2X^{N_2}$$

with analogous properties. The hypothesis about projective zeros covers both coordinate charts, so both applications of the Euclidean algorithm are justified.

Take a primitive integer pair (m,n), and let

$$\delta=\gcd\big(F(m,n),G(m,n)\big)>0.$$

Evaluation of the two identities gives

$$\delta\mid D_1n^{N_1},\qquad \delta\mid D_2m^{N_2},$$
$$\delta\mid \gcd(D_1D_2n^{N_1},D_1D_2m^{N_2})=|D_1D_2|,$$

where the last equality uses $\gcd(m,n)=1$. Consequently

$$1\le\delta\le D,\qquad D=|D_1D_2|.$$

This includes the cases $m=0$ or $n=0$, because the other coordinate is then 1 or -1.

Step 2: polynomial values have the expected size. Put $M=\max\{|m|,|n|\}$. The triangle inequality, applied to the coefficients of F and G, gives a constant $C\ge1$ with

$$\max\{|F(m,n)|,|G(m,n)|\}\le CM^d.$$

For the reverse estimate, consider the compact set

$$K=\{(s,t)\in\mathbb{R}^2:\max\{|s|,|t|\}=1\}.$$

The continuous function $\max\{|F(s,t)|,|G(s,t)|\}$ is positive on K. Compactness gives

$$\kappa=\min_{(s,t)\in K}\max\{|F(s,t)|,|G(s,t)|\}>0.$$

Since $(m/M,n/M)\in K$, homogeneity gives

$$\begin{aligned}\max\{|F(m,n)|,|G(m,n)|\}&=M^d\max\{|F(m/M,n/M)|,|G(m/M,n/M)|\}\\&\ge\kappa M^d.\end{aligned}$$

Step 3: reduce the output coordinates. By the definition of projective height,

$$H(\varphi([m:n])) = \frac{\max\{|F(m,n)|, |G(m,n)|\}}{\delta}.$$

The preceding bounds yield

$$\frac{\kappa}{D} M^d \le H(\varphi([m:n])) \le C M^d.$$

Taking logarithms proves the claim, for example with $c = \max\{0, \log C, \log(D/\kappa)\}$. □

24.6.2 Doubling increases height by a factor of four

Proposition 24.19 (The doubling estimate)**.** *There is a constant $c_0 \ge 0$, depending only on E, such that*

$$4h_E(P) - c_0 \le h_E(2P) \le 4h_E(P) + c_0 \qquad (P \in E(\mathbb{Q})).$$

Proof. For $P = (x,y)$ with $y \neq 0$, the tangent formula gives

$$\begin{aligned} x(2P) &= \left(\frac{3x^2 + a}{2y}\right)^2 - 2x \\ &= \frac{x^4 - 2ax^2 - 8bx + a^2}{4(x^3 + ax + b)}. \end{aligned}$$

The last equality uses $y^2 = x^3 + ax + b$ and expands the numerator. Define the homogeneous quartics

$$\begin{aligned} F(X,Z) &= X^4 - 2aX^2Z^2 - 8bXZ^3 + a^2Z^4, \\ G(X,Z) &= 4Z(X^3 + aXZ^2 + bZ^3). \end{aligned}$$

We check the hypothesis of Lemma 24.18. At $[1:0]$, the value of F is 1. Any other common projective zero would have the form $[t:1]$ with $f(t) = t^3 + at + b = 0$. But

$$F(t,1) = f'(t)^2 - 8tf(t) = f'(t)^2.$$

The polynomial f has distinct roots because $4a^3 + 27b^2 \neq 0$, so $f'(t) \neq 0$ at a root of f. There is no common zero.

Thus the quartics define a map φ on the projective line, and the duplication formula says $x(2P) = \varphi(x(P))$. This is also true at O, since $\varphi([1:0]) = [1:0]$. If $y = 0$, then $2P = O$ and the calculation $G(x,1) = 0$, $F(x,1) \neq 0$ gives the same conclusion. Apply Lemma 24.18 with $d = 4$. □

The lower estimate is the essential one. It says that doubling cannot turn points of arbitrarily large height into points of small height. The bounded cancellation established in the lemma is what makes this statement possible.

24.6.3 Translation by a fixed point

Proposition 24.20 (The translation estimate)**.** *For each fixed $R \in E(\mathbb{Q})$ there is a constant $c_R \geq 0$ such that*

$$h_E(P-R) \leq 2h_E(P) + c_R \qquad (P \in E(\mathbb{Q})).$$

Proof. For $R = O$ this holds with $c_R = 0$, since heights are nonnegative. Suppose $R = (u,v)$ is affine. We first consider $P = (x,y)$ with $x \neq u$. The line through P and $-R = (u,-v)$ has slope $(y+v)/(x-u)$. Hence

$$\begin{aligned} x(P-R) &= \frac{(y+v)^2}{(x-u)^2} - x - u \\ &= \frac{ux^2 + (u^2+a)x + au + 2b + 2vy}{(x-u)^2}. \end{aligned}$$

For the second equality, substitute $y^2 = x^3 + ax + b$ and $v^2 = u^3 + au + b$, and expand $(x+u)(x-u)^2$.

Write $x = A/C^2$ and $y = B/C^3$ as in the denominator proposition, and put

$$M = \max\{|A|, C^2\} = \exp(h_E(P)).$$

The denominator is reduced because $\gcd(A,C) = 1$. The equation gives

$$|B|^2 = |A^3 + aAC^4 + bC^6| \leq (1 + |a| + |b|)M^3.$$

Thus $|B| \leq c_E M^{3/2}$, where $c_E = (1+|a|+|b|)^{1/2}$, and $|BC| \leq c_E M^2$. After multiplying the preceding expression for $x(P-R)$ by C^4/C^4, it becomes

$$\frac{uA^2 + (u^2+a)AC^2 + (au+2b)C^4 + 2vBC}{(A - uC^2)^2}.$$

Every displayed term in its numerator and denominator is bounded in absolute value by a constant, depending on E, R, times M^2.

For precision, choose a positive integer L clearing the rational coefficients in both expressions. Then this fraction is N/D for integers N, D, with $D \neq 0$, and

$$\begin{aligned} |N| &\leq L\big(|u| + |u^2+a| + |au+2b| + 2|v|c_E\big)M^2, \\ |D| &\leq L(1+|u|)^2 M^2. \end{aligned}$$

Let $C_R \geq 1$ be at least both coefficient bounds. Reducing an integer fraction only decreases the maximum absolute value of its numerator and denominator. Therefore

$$H(x(P-R)) \leq C_R M^2, \qquad h_E(P-R) \leq 2h_E(P) + \log C_R.$$

Only the points $P = O$, $P = R$, and $P = -R$ were excluded; the last two are exactly the possible affine points with $x = u$. This is a finite set. Enlarge $\log C_R$, if necessary, to

dominate $h_E(P-R)-2h_E(P)$ for those points and to be nonnegative. The resulting constant c_R works for every P. □

24.6.4 The final descent

We have now proved every arithmetic ingredient needed for finite generation: weak Mordell gives finitely many classes modulo doubling, the doubling estimate gives growth of height, and the translation estimate controls subtraction of a fixed representative. For another treatment of height and descent, see [5, Chapter IV, §4].

Proof of Theorem 24.3. By the weak Mordell theorem, choose finitely many representatives $R_1,\dots,R_s$ for $E(\mathbb{Q})/2E(\mathbb{Q})$. Thus every $P\in E(\mathbb{Q})$ has the form

$$P = R_i + 2Q$$

for some i and some $Q\in E(\mathbb{Q})$. Apply the two height estimates to this equality:

$$4h_E(Q) - c_0 \le h_E(2Q) = h_E(P-R_i) \le 2h_E(P) + c_{R_i}.$$

If

$$C = \frac{c_0 + \max_i c_{R_i}}{4},$$

then

$$h_E(Q) \le \frac{1}{2}h_E(P) + C.$$

In particular, $h_E(Q) < h_E(P)$ whenever $h_E(P) > 2C$.

Choose a real number

$$B \ge \max\{2C, h_E(R_1),\dots,h_E(R_s)\}.$$

The set $S = \{P\in E(\mathbb{Q}) : h_E(P)\le B\}$ is finite, by the bounded-height argument, and it contains all the R_i. Let A be the subgroup generated by S.

Suppose that some point lies outside A. There is then a point $P\notin A$ of least height. To justify the minimum without assuming that heights are integers, start with any $P_0\notin A$. There are only finitely many points with height at most $h_E(P_0)$; choose a least-height point outside A among them. A point of still smaller height anywhere would also belong to that finite set, so this is a global minimum outside A.

Since $S\subseteq A$, we have $h_E(P) > B \ge 2C$. Write $P = R_i + 2Q$. If $Q\in A$, then $R_i\in A$ gives $P\in A$, which is impossible. Thus $Q\notin A$. But the preceding estimate gives $h_E(Q) < h_E(P)$, contradicting the choice of P.

It follows that $A = E(\mathbb{Q})$. The finite set S generates $E(\mathbb{Q})$, proving Mordell's theorem. The structure theorem for finitely generated abelian groups now gives $E(\mathbb{Q}) \cong \mathbb{Z}^r \oplus T$, with T finite. □

The proof explains why the word *descent* is appropriate: any point outside the subgroup generated by sufficiently small points would produce a smaller point outside the same subgroup. It also explains why the finiteness of $E(\mathbb{Q})/2E(\mathbb{Q})$ alone would not be enough. The height estimates are what force this descent to stop.

Making the descent termination quantitative

The recurrence in the proof can be iterated explicitly. If $P_j = R_{i_j} + 2P_{j+1}$ and $h_j = h_E(P_j)$, then

$$h_{j+1} \le \tfrac{1}{2}h_j + C \quad \Longrightarrow \quad h_n \le 2^{-n}h_0 + 2C(1-2^{-n}).$$

The implication follows by induction: multiply the expression for h_n by $1/2$ and add C. In particular, once $2^{-n}h_0 \le 1$, one has $h_n \le 2C+1$. Expanding the group identities gives

$$P_0 = R_{i_0} + 2R_{i_1} + \cdots + 2^{n-1}R_{i_{n-1}} + 2^n P_n.$$

All representatives and all possible terminal points P_n belong to one finite set after enlarging the height bound to contain the representatives. This provides a second proof of finite generation, without selecting a minimal-height counterexample.

The argument is an existence proof of a finite generating set. To turn it into an effective computation, one needs explicit constants in the height inequalities and a way to determine suitable representatives modulo two. The fact that finitely many classes occur does not by itself list them. This distinction concerns the content of the proof, not the truth of finite generation.

A canonical height from the doubling estimate

The two-sided estimate already proved also yields a useful limit, with no new arithmetic theorem. Suppose

$$|h_E(2P) - 4h_E(P)| \le c_0.$$

For fixed P, set $a_n = 4^{-n}h_E(2^nP)$. Then

$$|a_{n+1} - a_n| \le c_0\, 4^{-(n+1)}.$$

The geometric series converges, so (a_n) is Cauchy. Define the normalized height

$$\widehat{h}_x(P) = \lim_{n\to\infty} 4^{-n}h_E(2^nP).$$

Summing the error bound and shifting the index show

$$|\widehat{h}_x(P) - h_E(P)| \le c_0/3, \qquad \widehat{h}_x(2P) = 4\widehat{h}_x(P), \qquad \widehat{h}_x(P) \ge 0.$$

This convention uses the full x-coordinate height; another common convention divides it by two.

If P is torsion, its iterated doubles range over finitely many points, so $\widehat{h}_x(P) = 0$. Conversely, if $\widehat{h}_x(P) = 0$, the doubling identity gives $\widehat{h}_x(2^n P) = 0$ for all n, and the error bound yields $h_E(2^n P) \le c_0/3$. Bounded-height finiteness forces $2^r P = 2^s P$ for some $r < s$, so $(2^s - 2^r)P = O$. Thus

$$\widehat{h}_x(P) = 0 \quad \Longleftrightarrow \quad P \text{ is torsion.}$$

Only the doubling identity has been proved here. A quadratic addition law for this height is a further statement and is not required for the descent proof.

24.7 Counting points over a finite field

For a prime $p \ge 5$, reduction modulo p means applying the coefficient map $\mathbb{Z} \to \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ to an equation with integer coefficients. We may then seek solutions in the finite field $\mathbb{F}_p$. For an equation with integral coefficients, the reduced cubic is elliptic when its discriminant is nonzero modulo p; such a prime is called a *prime of good reduction for this model*. For $y^2 = x^3 - 2$, $\Delta = -1728$, so 5 and 7 are good primes.

In $\mathbb{F}_5$, the squares are $0, 1, 4$. Direct enumeration gives

x	0	1	2	3	4
$x^3 - 2$	3	4	1	0	2
$\#\{y : y^2 = x^3 - 2\}$	0	2	2	1	0

There are five affine points and the point at infinity, so $\#E(\mathbb{F}_5) = 6$. Here $P = (3,5)$ reduces to $(3,0)$, a point of order two. In the formula for $2P$, the denominator 100 is divisible by 5; one must not try to reduce that fraction by inverting 100. To reduce a rational projective point, choose primitive integer coordinates and reduce each coordinate modulo p. They cannot all become zero because their greatest common divisor is one. This procedure gives the correct result: $2P = [1290 : -383 : 1000]$ reduces to $[0 : 2 : 0] = O$.

The point-count exercise at the end of this lecture compares reduction modulo 7 with the computation modulo 5. Comparing reductions can constrain rational torsion, but that conclusion requires a theorem about reduction on torsion; a point count by itself is not a proof of infinite order. The examples already demonstrate the central connection of this lecture: elementary field arithmetic produces a finite group, while the corresponding rational group is governed by a deep finiteness theorem.

24.8 Exercises

Exercise 24.1 (Verify the answer by substitution)**.** Check directly that the point

$$\left(\frac{129}{100}, -\frac{383}{1000}\right)$$

lies on $y^2 = x^3 - 2$.

Exercise 24.2 (A finite search with a precise bound)**.** For $t = r/s \in \mathbb{Q}$ in lowest terms with $s > 0$, put $H(t) = \max(|r|, s)$. List all t with $H(t) \leq 2$. Let $E : y^2 = x^3 + ax + b$ be a fixed nonsingular curve with $a, b \in \mathbb{Z}$, and put $h_E(P) = \log H(x(P))$ for $P \neq O$ and $h_E(O) = 0$. Explain why the same method gives a finite search for $P \in E(\mathbb{Q})$ with $h_E(P) \leq \log 2$.

Exercise 24.3 (Finite quotients without finite generation)**.** Prove directly that the additive group $\mathbb{Q}$ is not finitely generated, even though every rational number is twice another rational number.

Exercise 24.4 (Square classes over the rational numbers)**.** Let $S = \{p_1, \ldots, p_s\}$ consist of s distinct rational primes. Describe all the classes in $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ represented by nonzero rational numbers u for which $v_p(u)$ is even at every prime $p \notin S$. Prove that there are exactly 2^{s+1} such classes, and list representatives for $S = \{2, 3, 5\}$.

Exercise 24.5 (A concrete obstruction to halving)**.** On $E : y^2 = x^3 - x$ use the ordered roots $(e_1, e_2, e_3) = (0, 1, -1)$. Compute the descent triples $\delta(P)$ for $P = (0,0), (1,0), (-1,0)$. Deduce that none of these three points is twice a rational point of E.

Exercise 24.6 (A second point count)**.** Count the points over $\mathbb{F}_7$ of the projective elliptic curve with affine equation $y^2 = x^3 - 2$, including its point at infinity.

Exercise 24.7 (Arithmetic on a cubic)**.** On the projective elliptic curve with affine equation $E : y^2 = x^3 - 2$ over $\mathbb{Q}$, compute $2P$ for $P = (3, 5)$. Explain exactly what this computation and Mordell's theorem say about $E(\mathbb{Q})$.

Lecture 25
Sheaves: local information and gluing

A sheaf records sections on open sets and the rules for restricting and gluing them. The structure sheaf puts the local rings of a variety or a spectrum into a single object. We construct it from fractions and prove that exactness of sheaves is detected on stalks. References are [2, Chapter II, §§1–2]; localization from [1, Chapter 3] remains the main algebraic tool.

25.1 Presheaves and sheaves

Suppose an open set U is covered by open subsets U_i. A continuous function on U restricts to a continuous function on every U_i. Conversely, continuous functions s_i on the U_i determine a function on U if their values agree wherever two domains overlap. The glued function is unique, and continuity can be checked in a neighborhood of each point. A sheaf isolates this pattern, without requiring its sections to be ordinary functions.

Definition 25.1. A *presheaf of abelian groups* $\mathscr{F}$ on a topological space X assigns an abelian group $\mathscr{F}(U)$ to each open set U, and a restriction homomorphism

$$\rho_{UV} : \mathscr{F}(U) \longrightarrow \mathscr{F}(V) \quad (V \subseteq U),$$

with $\rho_{UU} = \mathrm{id}$ and $\rho_{VW}\rho_{UV} = \rho_{UW}$ when $W \subseteq V \subseteq U$. We write $s|_V$ for $\rho_{UV}(s)$ and call s a *section*. We take $\mathscr{F}(\varnothing) = 0$.

It is a *sheaf* if, for every open cover $U = \bigcup_i U_i$, compatible sections $s_i \in \mathscr{F}(U_i)$, meaning

$$s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j},$$

are restrictions of exactly one section $s \in \mathscr{F}(U)$.

The words “exactly one” combine existence of gluing with local determination. In particular a section whose restrictions to a cover are zero is zero. A *sheaf of rings* assigns rings to opens and ring homomorphisms to inclusions, with the same gluing axiom. Its value on the empty set is the zero ring. Given a sheaf of rings $\mathscr{O}_X$, an *$\mathscr{O}_X$-module* is a sheaf $\mathscr{F}$ of abelian groups such that $\mathscr{F}(U)$ is an $\mathscr{O}_X(U)$-module for every open U, and

$$(as)|_V = (a|_V)(s|_V) \qquad (V \subseteq U).$$

Thus multiplication of a section by a local scalar commutes with restriction.

We write $\Gamma(U, \mathscr{F}) = \mathscr{F}(U)$. A section on the whole space X is a *global section*. For an open subset $U \subseteq X$, the *restriction sheaf* $\mathscr{F}|_U$ is defined by $(\mathscr{F}|_U)(V) = \mathscr{F}(V)$ for each open $V \subseteq U$.

Example 25.2. For a topological space X, $\mathscr{C}_X(U)$ is the ring of continuous functions $U \to \mathbb{R}$. Compatible sections glue point by point, and the result is continuous because continuity is local. On an algebraic variety, regular functions form the structure sheaf $\mathscr{O}_X$: locally they are quotients of polynomials with nonvanishing denominator. Gluing preserves this local property.

The constant presheaf and its sheafification

Let V be an abelian group. Assigning the group V to each nonempty open set, with identity restrictions, defines the constant presheaf. It need not be a sheaf. If $U = U_1 \sqcup U_2$ is a disjoint union of two nonempty opens, choose different elements $v_1, v_2 \in V$. They agree on the empty overlap, but no single element of V restricts to both. Gluing fails.

The constant sheaf instead assigns locally constant functions $U \to V$, where V is given the discrete topology. Such a function may take different values on different open pieces. Compatible locally constant functions glue pointwise, and the resulting function is locally constant near each point because it agrees there with one member of the cover. This verifies the sheaf axiom. On a connected open set every locally constant function is constant: the inverse images of a value and its complement would otherwise give a separation.

The stalk of the constant sheaf is V at every point, since a locally constant function is constant on some neighborhood of that point. Thus stalks do not tell us that global sections must be the same group on all opens; topology governs how local data fit together.

25.2 Germs and stalks

Often the domain of a section is less important than what happens sufficiently near one point.

Definition 25.3. Fix $p \in X$. Two pairs (U, s) and (V, t), with $p \in U \cap V$, define the same *germ at* p if there is an open neighborhood $W \subseteq U \cap V$ of p on which $s|_W = t|_W$. The set of these equivalence classes is the *stalk* $\mathscr{F}_p$. Addition is performed after restricting two representatives to their common domain. We write s_p for the germ of s.

The construction is an example of a *direct limit*, written $\varinjlim_{p \in U} \mathscr{F}(U)$: elements represented at different stages are identified when their images agree at a common later stage, here a smaller neighborhood. We use the explicit equivalence relation above, so no general

theory of direct limits is required. Equality of germs means equality on some neighborhood; it does not mean equality merely at the point.

Example 25.4. Over an algebraically closed field k, the regular functions 0 and x on $\mathbb{A}^1_k$ have the same value at 0, but different germs there. Their difference is not zero on any open neighborhood of 0. The stalk is $k[x]_{(x)}$, while the value at the point is obtained by quotienting by its maximal ideal:

$$k[x]_{(x)} \longrightarrow k[x]_{(x)}/(x) \cong k.$$

Thus a stalk remembers local behavior, and its residue field remembers values.

Proposition 25.5. *If $\mathscr{F}$ is a sheaf and $s,t \in \mathscr{F}(U)$, then $s = t$ if and only if $s_p = t_p$ for every $p \in U$.*

Proof. Only the reverse implication needs proof. For each p, equality of germs gives a neighborhood $U_p \subseteq U$ on which s and t agree. These neighborhoods cover U, and uniqueness in the gluing axiom gives $s = t$. □

25.3 The structure sheaf of a spectrum

Let $X = \operatorname{Spec} A$. We already know its basic open sets

$$D(f) = \{\mathfrak{p} \in \operatorname{Spec} A : f \notin \mathfrak{p}\}, \qquad D(f) \cap D(g) = D(fg).$$

Our goal is to make the assignment

$$\mathscr{O}_X(D(f)) = A_f$$

into a sheaf. If $D(g) \subseteq D(f)$, then f lies in no prime of A_g, and so is a unit in A_g. The universal property gives the restriction map $A_f \to A_g$. This works even if A has zero divisors or nilpotents.

We need to prove that fractions glue. The following algebraic argument is useful enough to state for modules.

Lemma 25.6 (Gluing fractions on a finite basic cover)**.** *Suppose $D(f_1), \ldots, D(f_r)$ cover* $\operatorname{Spec} A$*, and M is an A-module. If $s_i \in M_{f_i}$ have equal images in $M_{f_i f_j}$ for every i, j, there is a unique $s \in M$ restricting to all s_i.*

Proof. The cover condition means $(f_1, \ldots, f_r) = A$: otherwise a maximal ideal containing this ideal would be a point omitted by the cover. The same is true of $(f_1^N, \ldots, f_r^N)$ for every positive integer N, since taking powers does not change which prime ideals contain these elements.

For uniqueness, let s have zero image in every M_{f_i}. Some power $f_i^{n_i}$ kills s, by the definition of localization. Choose $N \geq n_i$ for every i and coefficients c_i with $\sum_i c_i f_i^N = 1$. Multiplying by s gives $s = 0$.

For existence, first choose a common n and $m_i \in M$ with $s_i = m_i/f_i^n$. Fix i, j. In M_{f_j}, the element $m_i - f_i^n s_j$ becomes zero after inverting f_i, because $s_i = s_j$ on the overlap. Thus some power of f_i kills it in M_{f_j}. There are finitely many pairs, so choose a common $N \geq n$ large enough that

$$f_i^N s_j = f_i^{N-n} m_i \quad \text{in } M_{f_j} \quad \text{for every } i, j.$$

Choose c_i with $\sum_i c_i f_i^N = 1$, and set

$$n_i = f_i^{N-n} m_i \in M, \qquad s = \sum_i c_i n_i.$$

Its image in each M_{f_j} is

$$s|_{D(f_j)} = \sum_i c_i n_i = \sum_i c_i f_i^N s_j = \left(\sum_i c_i f_i^N \right) s_j = s_j.$$

□

Every basic open $D(f) = \operatorname{Spec} A_f$ is quasi-compact. Thus an arbitrary cover by basic opens has a finite subcover. Apply the lemma to A_f to prove gluing on any basic open; uniqueness also verifies agreement with the remaining members of the cover. Since the basic opens form a basis closed under finite intersections, this defines a sheaf on all opens: a section over an arbitrary U is a compatible family of sections on basic opens contained in U.

Equivalently, it is a family $s(\mathfrak{p}) \in A_{\mathfrak{p}}$ locally represented by one fraction. This construction gives

$$\Gamma(\operatorname{Spec} A, \mathscr{O}_X) = A, \qquad \mathscr{O}_{X,\mathfrak{p}} = A_{\mathfrak{p}}.$$

For the stalk identity, every fraction $a/s \in A_{\mathfrak{p}}$ is represented on $D(s)$. If two fractions become equal in $A_{\mathfrak{p}}$, some element outside $\mathfrak{p}$ kills the cross-multiplied difference. Restricting to the corresponding smaller basic neighborhood makes them equal. This proves both surjectivity and injectivity of the natural map from the stalk.

Definition 25.7. A *ringed space* is a topological space X together with a sheaf of rings $\mathscr{O}_X$. It is *locally ringed* if every stalk $\mathscr{O}_{X,p}$ is a local ring. An *affine scheme* is the locally ringed space $(\operatorname{Spec} A, \mathscr{O}_{\operatorname{Spec} A})$ just constructed from a ring A. A *scheme* is a locally ringed space covered by open subsets isomorphic to affine schemes.

An isomorphism here includes both a homeomorphism of spaces and ring isomorphisms on corresponding open sets that commute with restriction. An open set equipped with one of these affine descriptions is called an *affine open*. For example $\operatorname{Spec} k[x]$ and $\operatorname{Spec} \mathbb{Z}$ are affine schemes. The scheme $\operatorname{Spec} k[\varepsilon]/(\varepsilon^2)$ has one point, but its ring of sections contains the nonzero nilpotent ε. This example shows why the underlying point set does not determine the scheme.

The projective line is a scheme covered by two affine lines: on the overlap their coordinates satisfy $s = t^{-1}$, and the corresponding rings of sections are both $k[t, t^{-1}]$. This gluing identifies the structure sheaves as well as the open subsets. It extends the affine-chart description of projective space from Lecture 19.

Gluing fractions in a concrete two-open cover

On $\operatorname{Spec}\mathbb{Z}$, the two basic opens $D(2)$ and $D(3)$ cover because $(2,3)=\mathbb{Z}$. Their intersection is $D(6)$. A pair

$$a \in \mathbb{Z}[1/2], \qquad b \in \mathbb{Z}[1/3]$$

glues precisely when it gives the same element of $\mathbb{Z}[1/6]$. All these rings embed in $\mathbb{Q}$, so this is equality of rational numbers. Such a common number must be an integer: in a reduced fraction its denominator is simultaneously a power of two and a power of three, and hence is one. Thus

$$\mathbb{Z} = \mathbb{Z}[1/2] \cap \mathbb{Z}[1/3] \subseteq \mathbb{Z}[1/6].$$

The pair $(1/2, 1/3)$ is not compatible and cannot be glued. A cover does not mean that arbitrary local values produce a section; agreement on the overlap is necessary.

For a general ring with zero divisors, the localizations need not embed in a common fraction field. The gluing lemma therefore uses annihilating powers rather than an intersection of subrings. Its identity $\sum_i c_i f_i^N = 1$ plays the role of combining compatible local data into one global element. This is why the fraction proof was formulated for arbitrary modules.

25.4 Skyscraper sheaves and morphisms

For a closed point $p \in X$ and an abelian group V, the *skyscraper sheaf* i_*V, where $i : \{p\} \hookrightarrow X$, is

$$(i_*V)(U) = \begin{cases} V, & p \in U, \\ 0, & p \notin U. \end{cases}$$

Restriction is the identity when both opens contain p and the zero map otherwise. If a cover contains p, at least one member contains p, and compatible values on all such members are the same; this proves gluing. Its stalk is V at p and zero elsewhere. The closedness hypothesis ensures that every other point has a neighborhood avoiding p.

Definition 25.8. A *morphism of sheaves* $\varphi : \mathscr{F} \to \mathscr{G}$ is a collection of homomorphisms $\mathscr{F}(U) \to \mathscr{G}(U)$ commuting with restrictions. It induces a homomorphism $\varphi_p : \mathscr{F}_p \to \mathscr{G}_p$ by sending the germ of s to the germ of $\varphi(s)$. For sheaves of $\mathscr{O}_X$-modules we require each map on sections to be $\mathscr{O}_X(U)$-linear. An *isomorphism* of sheaves is a morphism with an inverse morphism.

Proposition 25.9 (Isomorphisms are detected on stalks)**.** *A morphism of sheaves is an isomorphism if and only if all its stalk maps are isomorphisms.*

Proof. Suppose all stalk maps are isomorphisms. If $\varphi(s) = 0$, injectivity at every stalk gives $s_p = 0$ everywhere; hence $s = 0$. Thus the maps on every open set are injective.

Given $t \in \mathscr{G}(U)$, stalkwise surjectivity gives, for each $p \in U$, a germ in $\mathscr{F}_p$ lifting t_p. Represent it by a local section and shrink its neighborhood until its image is exactly t. On an overlap two such lifts have equal images, and the injectivity just proved makes the lifts equal. They therefore glue to a section mapping to t. Thus every open-set map is bijective;

the inverses commute with restrictions because the original maps do. The converse is immediate. □

25.5 Kernels, cokernels, and exactness

The kernel of $\varphi : \mathscr{F} \to \mathscr{G}$ is given on opens by $\ker \varphi(U)$. This is a sheaf: local sections in the kernel glue in $\mathscr{F}$, and their image is locally zero, hence zero. In contrast, the presheaf $U \mapsto \mathscr{G}(U)/\varphi(\mathscr{F}(U))$ may fail the gluing axiom.

Its repair is *sheafification*. For a presheaf $\mathscr{H}$, form families of germs $s(p) \in \mathscr{H}_p$ which, near each point, are represented by a section of $\mathscr{H}$. Such families glue by their values and form a sheaf $\mathscr{H}^+$. Its stalks are the original $\mathscr{H}_p$: local representatives show surjectivity, and equality of germs after shrinking shows injectivity. The *cokernel sheaf* of φ is the sheafification of the presheaf $U \mapsto \mathscr{G}(U)/\varphi(\mathscr{F}(U))$. The *image sheaf* is the sheafification of $U \mapsto \varphi(\mathscr{F}(U))$; it is naturally a subsheaf of $\mathscr{G}$. A sequence $\mathscr{F} \xrightarrow{\varphi} \mathscr{G} \xrightarrow{\psi} \mathscr{H}$ is *exact at* $\mathscr{G}$ if $\operatorname{im}\varphi = \ker\psi$ as subsheaves. A morphism is called injective if its kernel is zero, and surjective if its cokernel sheaf is zero. These definitions concern sheaves, so surjectivity has not been defined to mean surjectivity on each open set.

Proposition 25.10 (Stalk criterion for exactness)**.** *A sequence of sheaves is exact if and only if its sequence of stalks is exact at every point. In particular, a morphism is surjective as a sheaf map if and only if sections of its target lift locally near every point.*

Proof. Stalks commute with kernels: if the image germ of s is zero, shrink until the image section is zero. Stalks also commute with cokernels: a germ of a section is zero in the cokernel stalk precisely when, after shrinking, it is the image of a local section. Sheafification does not change stalks, as explained above. Thus the stalk of the quotient of a kernel by an image is the corresponding kernel modulo image of stalks. A sheaf with every stalk zero is zero, by local determination. These observations prove the equivalence. □

Local surjectivity need not imply surjectivity on global sections; the logarithm exercise at the end of this lecture gives a concrete example. In the next lecture we construct a cochain complex whose degree-one cohomology measures a discrepancy between local sections. This gives a concrete geometric application of the complexes and cohomology groups defined in Lecture 14.

A surjective sheaf map which is not onto on global sections

Here is an explicit example of the local-to-global distinction. On the topological circle S^1, let $\mathscr{C}_{\mathbb{R}}$ be the sheaf of continuous real-valued functions and $\mathscr{C}_{S^1}$ the sheaf of continuous functions valued in the multiplicative group $S^1 \subset \mathbb{C}$. The map

$$\mathscr{C}_{\mathbb{R}} \longrightarrow \mathscr{C}_{S^1}, \qquad f \longmapsto e^{2\pi i f}$$

is surjective as a map of sheaves: near a fixed point the values of a continuous circle-valued function lie in a sufficiently short arc, on which a continuous argument can be chosen. Its kernel is the sheaf of locally constant integer-valued functions.

The global section $g(z) = z$ has no continuous real lift. If $g(z) = e^{2\pi i f(z)}$, define $F(t) = f(e^{2\pi i t})$ for $t \in [0,1]$. Then $F(t) - t$ is continuous and integer-valued, so it is constant. Hence $F(1) - F(0) = 1$. But the two endpoint inputs to f are the same point, so $F(1) = F(0)$, a contradiction.

The resulting sequence of sheaves is exact, but global sections do not preserve its final surjection. A long exact cohomology sequence records the obstruction by a connecting homomorphism. This example should be compared with Lecture 14: local lifts take the place of degreewise lifts, and their discrepancies produce the next cohomology class.

25.6 Exercises

Exercise 25.1 (Constant presheaf versus constant sheaf)**.** Let $X = \{p, q\}$ with the discrete topology. Assign $\mathbb{Z}$ to every nonempty open set, assign the zero group to the empty set, and use identity restriction maps between nonempty sets and the unique maps to the zero group for restriction to the empty set. Show that this presheaf is not a sheaf. Repair it.

Exercise 25.2 (Gluing that can be checked by hand)**.** Cover $\operatorname{Spec}\mathbb{Z}$ by $D(2)$ and $D(3)$. Explain why a pair $a \in \mathbb{Z}[1/2]$, $b \in \mathbb{Z}[1/3]$ agreeing in $\mathbb{Z}[1/6]$ comes from a unique integer.

Exercise 25.3 (Local lifts need not give a global lift)**.** On the complex plane $\mathbb{C}$ with its usual topology, let $\mathscr{C}$ be the sheaf of continuous complex-valued functions under pointwise addition, and $\mathscr{C}^\times$ the sheaf of continuous nowhere-zero complex-valued functions under pointwise multiplication. Explain why $s \mapsto e^{2\pi i s}$ is a surjective morphism of sheaves of abelian groups but is not surjective on sections over $\mathbb{C}^\times$.

Lecture 26
Modules become sheaves; vector bundles on the projective line

Modules determine sheaves on affine schemes by localization. Gluing free modules across affine charts then produces vector bundles, and an explicit calculation on the projective line describes their sections and a gluing obstruction. References are [2, Chapter II, §5], and Chapter III, §§4–5 for the last section.

26.1 The sheaf associated to a module

Let $X = \operatorname{Spec} A$ and let M be an A-module. The gluing lemma from Lecture 25 applies to M, not only to A. It therefore constructs a sheaf $\widetilde{M}$ by

$$\widetilde{M}(D(f)) = M_f.$$

Multiplication $A_f \times M_f \to M_f$ is compatible with restriction. Thus $\widetilde{M}$ is an $\mathscr{O}_X$-module in the sense defined in Lecture 25: local functions act on its sections and this action commutes with restriction.

Exactly as for the structure sheaf,

$$(\widetilde{M})_{\mathfrak{p}} = M_{\mathfrak{p}}, \qquad \Gamma(X, \widetilde{M}) = M.$$

These identities distinguish the whole module, which describes global sections on an affine scheme, from its localizations, which describe nearby sections or germs.

Example 26.1. Take $A = k[x]$ and $M = A/(x)$. If a basic open $D(f)$ contains the point (x), then $f(0) \neq 0$ and $M_f \cong k$. If it does not contain (x), then $f \in (x)$ acts as zero on M, and inverting it gives $M_f = 0$. Hence $\widetilde{A/(x)}$ is the skyscraper sheaf k at the origin. In contrast, $\widetilde{A} = \mathscr{O}_X$ has nonzero stalks everywhere.

Proposition 26.2. *The operation $M \mapsto \widetilde{M}$ preserves exact sequences, and*

$$\operatorname{Hom}_A(M,N) \cong \operatorname{Hom}_{\mathscr{O}_X}(\widetilde{M}, \widetilde{N}).$$

Proof. An A-linear map $\alpha : M \to N$ gives compatible maps

$$\alpha_f : M_f \longrightarrow N_f, \qquad \frac{m}{f^n} \longmapsto \frac{\alpha(m)}{f^n},$$

hence a sheaf map $\widetilde{\alpha}$. Conversely, for a sheaf map φ, set

$$\alpha = \Gamma(X, \varphi) : M \longrightarrow N.$$

Restriction and $\mathscr{O}_X$-linearity force

$$\varphi_{D(f)}(m/f^n) = f^{-n}\varphi_{D(f)}(m/1) = \alpha(m)/f^n.$$

Thus $\varphi = \widetilde{\alpha}$ and $\Gamma(X, \widetilde{\alpha}) = \alpha$, proving the Hom identity.

Localization is exact, as proved in Lecture 7. Therefore an exact sequence of modules gives an exact sequence at every stalk. The stalk criterion gives exactness of the associated sheaves. □

Definition 26.3. An $\mathscr{O}_X$-module $\mathscr{E}$ is *locally free of rank r* if every point has an open neighborhood U with $\mathscr{E}|_U \cong \mathscr{O}_U^r$. Here $\mathscr{O}_U^r$ is the sheaf of r-tuples of local functions. Its standard basis consists of the tuples with one entry 1 and the remaining entries 0. A locally free sheaf of rank one is a *line bundle*, also called an invertible sheaf. The *support* of any sheaf $\mathscr{F}$ is

$$\operatorname{Supp} \mathscr{F} = \{p \in X : \mathscr{F}_p \neq 0\}.$$

26.2 Quasi-coherent and coherent sheaves

A scheme is *locally Noetherian* if it has an affine open cover $\operatorname{Spec} A_i$ with each A_i Noetherian. It is *Noetherian* if, in addition, its underlying topological space is quasi-compact; equivalently, one can choose a finite such affine cover. For instance, every affine scheme of a finitely generated k-algebra is Noetherian by the Hilbert basis theorem. The projective line is Noetherian because its two standard affine coordinate rings are polynomial rings in one variable.

Definition 26.4. An $\mathscr{O}_X$-module $\mathscr{F}$ on a scheme X is *quasi-coherent* if X has an affine open cover $U_i = \operatorname{Spec} A_i$ on which $\mathscr{F}|_{U_i} \cong \widetilde{M_i}$ for modules M_i. On a Noetherian scheme, it is *coherent* if these modules can be taken finitely generated.

We restrict the word coherent to Noetherian schemes here. Over a Noetherian ring finite generation also gives finite presentation, since the relations among finitely many generators form a submodule of a finite free module. Thus a coherent sheaf is locally described by finitely many generators and finitely many relations. There is a more careful definition beyond the Noetherian setting, which is unnecessary for our course.

Theorem 26.5 (Affine module–sheaf correspondence)**.** *For any ring A, the constructions*

$$M \longmapsto \widetilde{M}, \qquad \mathscr{F} \longmapsto \Gamma(\operatorname{Spec} A, \mathscr{F})$$

are inverse equivalences between A-modules and quasi-coherent sheaves on Spec$A$*. Here an* equivalence *means that the constructions recover objects up to compatible isomorphisms and give bijections on the sets of morphisms. If A is Noetherian, they restrict to finitely generated modules and coherent sheaves.*

This is [2, Chapter II, §5, Corollary 5.5]. We have proved the assertion about morphisms and $\Gamma(X, \widetilde{M}) = M$. It remains to show that a sheaf with module descriptions on small affine neighborhoods has a single module description on the whole affine scheme. Here are the details of the required denominator argument.

We first justify refining affine charts. Suppose $W = \operatorname{Spec} C$ is an affine open inside $V = \operatorname{Spec} B$, and N is a B-module. Then

$$(\widetilde{N})|_W \cong \widetilde{N \otimes_B C}.$$

Indeed restrictions of sections and multiplication by local functions give the comparison map. At a point $q \in W$, corresponding to $p \in V$, both stalks are $N \otimes_B B_p$, since $C_q \cong B_p$ is the local ring of the same point in the open subspace. The stalk criterion proves the comparison is an isomorphism.

Now let $\mathscr{F}$ be quasi-coherent on $X = \operatorname{Spec} A$ and set $M = \Gamma(X, \mathscr{F})$. Choose a finite basic cover $D(g_i)$ with $\mathscr{F}|_{D(g_i)} = \widetilde{N}_i$. Such a cover exists: refine the affine cover in the definition by basic opens of X, use the preceding comparison to retain module descriptions, and use quasi-compactness to make the cover finite.

Fix $f \in A$. If $s \in M$ restricts to zero on $D(f)$, its image in each N_i is killed by a power of f. Finitely many i allow a common power f^n, and the sheaf axiom gives $f^n s = 0$ globally. Thus the natural map

$$M_f \longrightarrow \mathscr{F}(D(f))$$

is injective. To prove surjectivity, let $t \in \mathscr{F}(D(f))$. On $D(fg_i)$ it is an element of $(N_i)_f$. Clearing the finitely many denominators gives sections t_i on $D(g_i)$ restricting to $f^n t$, with one common n. On $D(g_i g_j)$ the differences $t_i - t_j$ vanish after inverting f. Applying the just-proved vanishing argument on this affine open, and choosing one common exponent m for the finitely many pairs, makes the sections $f^m t_i$ agree. They glue to a global s with $s|_{D(f)} = f^{n+m} t$. Therefore t is the image of s/f^{n+m}.

We have proved $M_f = \mathscr{F}(D(f))$ on every basic open, compatibly with restrictions. Hence $\widetilde{M} \cong \mathscr{F}$. For the coherent assertion, suppose each M_{g_i} is finitely generated. Lift numerators of a finite set of generators from each localization and let $M' \subseteq M$ be the submodule generated by all those numerators. Then $(M/M')_{g_i} = 0$ for every i. The uniqueness part of the gluing lemma implies $M/M' = 0$, so M is finitely generated. This finishes the affine correspondence.

Remark 26.6. An exact sequence of quasi-coherent sheaves on an affine scheme gives an exact sequence on global sections: translate it into modules, and use exactness of the equivalence. The affine and quasi-coherent hypotheses matter. The local logarithm example in Lecture 25 showed why one cannot assert this for arbitrary sheaves.

Tensor products, stalks, and fibers of associated sheaves

For A-modules M,N, localization of tensor products gives compatible identifications on basic opens,

$$(M \otimes_A N)_f \cong M_f \otimes_{A_f} N_f.$$

They glue to $\widetilde{M \otimes_A N} \cong \widetilde{M} \otimes_{\mathscr{O}_X} \widetilde{N}$. At a point $\mathfrak{p}$,

$$\widetilde{M}(\mathfrak{p}) = M \otimes_A \kappa(\mathfrak{p}) = M_{\mathfrak{p}}/\mathfrak{p}M_{\mathfrak{p}}.$$

Passing to a stalk is exact because it is localization. Passing to a fiber includes tensoring with a residue field and need not be exact. For example, the inclusion of sheaves corresponding to $(t) \hookrightarrow k[t]$ becomes the zero map on fibers at $t = 0$:

$$(t)/(t^2) \longrightarrow k[t]/(t), \qquad [t] \longmapsto 0.$$

The source is one-dimensional. This is the same failure of tensor exactness encountered in Lecture 5, now expressed geometrically.

For a coherent sheaf on a Noetherian scheme, a zero fiber at p implies a zero stalk there by Nakayama. A zero stalk in an affine neighborhood, together with finite generation, implies vanishing on a smaller basic neighborhood: choose one denominator killing all generators. Thus the support of a coherent sheaf is closed. Without the finiteness assumption a zero fiber need not imply a zero stalk; the local-ring module $\mathbb{Q}$ over $\mathbb{Z}_{(p)}$ again gives a counterexample.

26.3 Locally free sheaves and their fibers

For a ring R, $\mathrm{GL}_r(R)$ denotes the group of invertible $r \times r$ matrices over R, and I_r is its identity matrix. A *vector bundle of rank r* is obtained by gluing copies of affine r-space over open sets U_i, with linear invertible changes of vector coordinates on overlaps. Over an affine chart $U_i = \mathrm{Spec} A_i$, the local product $U_i \times \mathbb{A}^r$ means $\mathrm{Spec} A_i[z_1, \ldots, z_r]$, with the map to U_i induced by the inclusion of constants. To identify vector coordinates v_i and v_j on an overlap, specify

$$v_j = g_{ji} v_i, \qquad g_{ji} \in \mathrm{GL}_r(\mathscr{O}_X(U_i \cap U_j)).$$

The *cocycle condition* is

$$g_{ii} = I_r, \qquad g_{kj} g_{ji} = g_{ki} \quad \text{on } U_i \cap U_j \cap U_k.$$

It ensures consistent changes of coordinates:

$$v_k = g_{kj} v_j = g_{kj} g_{ji} v_i = g_{ki} v_i, \qquad g_{ij} = g_{ji}^{-1}.$$

These identifications glue the local products into the total space of the vector bundle. A section chooses a vector over each base point, described on U_i by r regular functions

satisfying the displayed coordinate-change rule. Its sheaf of sections is therefore locally free of rank r. Conversely, choosing a basis for a locally free sheaf on each U_i gives these matrices and reconstructs the bundle. We shall work mainly with the sheaf description. A chosen local basis is also called a *frame*.

At a point p, the stalk $\mathscr{E}_p$ is a module over the local ring $\mathscr{O}_{X,p}$. Its *fiber* is the vector space

$$\mathscr{E}(p) = \mathscr{E}_p \otimes_{\mathscr{O}_{X,p}} \kappa(p) = \mathscr{E}_p/\mathfrak{m}_p\mathscr{E}_p,$$

where $\kappa(p) = \mathscr{O}_{X,p}/\mathfrak{m}_p$ is the residue field. For a locally free sheaf of rank r, this vector space has dimension r. A stalk contains functions near a point; a fiber retains only their values there.

26.4 Constructing $\mathscr{O}(n)$ on $\mathbb{P}^1$

Let $\mathbb{P}^1_k$ have homogeneous coordinates $[X_0 : X_1]$. Its standard cover is

$$U_0 = \{X_0 \neq 0\} = \operatorname{Spec} k[t], \quad t = X_1/X_0, \qquad U_1 = \{X_1 \neq 0\} = \operatorname{Spec} k[s], \quad s = X_0/X_1.$$

The overlap has $s = t^{-1}$ and coordinate ring $k[t,t^{-1}]$. For an integer n, take one free generator e_0 on U_0 and another e_1 on U_1, with transition relation

$$\boxed{e_1 = t^n e_0.}$$

Since t is a unit on the overlap, this is an invertible change of basis. The resulting line bundle is denoted $\mathscr{O}_{\mathbb{P}^1}(n)$. This equation fixes our sign convention completely.

A global section is a pair $p(t)e_0$, $q(s)e_1$, where $p \in k[t]$ and $q \in k[s]$, satisfying

$$p(t) = t^n q(t^{-1}).$$

Write $q(s) = \sum_{j\geq 0} b_j s^j$. Then

$$p(t) = \sum_{j\geq 0} b_j t^{n-j} \in k[t] \quad \Longleftrightarrow \quad b_j = 0 \text{ whenever } j > n.$$

Consequently

$$\Gamma(\mathbb{P}^1, \mathscr{O}(n)) = \begin{cases} \langle 1, t, \ldots, t^n \rangle_k e_0, & n \geq 0, \\ 0, & n < 0, \end{cases}$$

and its dimension is $\max\{n+1, 0\}$.

For $n \geq 0$, the correspondence with homogeneous polynomials is explicit. A polynomial $F(X_0, X_1)$ of degree n gives $p(t) = F(1,t)$ and $q(s) = F(s,1)$, and homogeneity says $F(1,t) = t^n F(t^{-1}, 1)$. Thus the basis above corresponds to $X_0^n, X_0^{n-1}X_1, \ldots, X_1^n$. In particular $\Gamma(\mathbb{P}^1, \mathscr{O}) = k$: every global regular function on the projective line is constant.

For $\mathscr{O}_X$-modules $\mathscr{F}, \mathscr{G}$, their *tensor product sheaf* $\mathscr{F} \otimes_{\mathscr{O}_X} \mathscr{G}$ is obtained by sheafifying the presheaf

$$U \longmapsto \mathscr{F}(U) \otimes_{\mathscr{O}_X(U)} \mathscr{G}(U).$$

For a locally free sheaf $\mathscr{E}$, its *dual sheaf* $\mathscr{E}^\vee$ assigns to U the $\mathscr{O}_U$-linear sheaf maps $\mathscr{E}|_U \to \mathscr{O}_U$. On a chart with a basis $e_1, \dots, e_r$, the dual basis $e_1^\vee, \dots, e_r^\vee$ is characterized by $e_i^\vee(e_j) = \delta_{ij}$, with $\delta_{ij} = 1$ for $i = j$ and 0 otherwise.

For line bundles these operations can be computed on frames. The tensor frame $e_i \otimes e_i'$ has transition factor $t^n t^m = t^{n+m}$, whereas the dual frame has transition factor t^{-n}. Therefore

$$\mathscr{O}(n) \otimes \mathscr{O}(m) \cong \mathscr{O}(n+m), \qquad \mathscr{O}(n)^\vee \cong \mathscr{O}(-n).$$

These identities follow from the transition relations and require no classification theorem for line bundles.

Computing sections from a transition function

Use $t = X_1/X_0$ on U_0 and $u = X_0/X_1 = t^{-1}$ on U_1. The convention for $\mathscr{O}(n)$ is $e_1 = t^n e_0$. A global section consists of $f_0(t)e_0$ and $f_1(u)e_1$ agreeing on the overlap, which means

$$f_0(t) = t^n f_1(t^{-1}).$$

For $n \geq 0$, if $f_0(t) = \sum_{j=0}^n a_j t^j$, then

$$f_1(u) = \sum_{j=0}^n a_j u^{n-j}$$

is a polynomial and gives the unique compatible section. Conversely the compatibility equation forces all exponents of f_0 to lie between zero and n. Hence

$$H^0(\mathbb{P}^1, \mathscr{O}(n)) \cong \langle 1, t, \dots, t^n \rangle_k \qquad (n \geq 0).$$

For $n < 0$, the left side of the compatibility equation has only nonnegative exponents of t, while the right side has only exponents at most $n < 0$. The only possibility is zero. For $n = 0$ this also proves that every global regular function on $\mathbb{P}^1$ is constant.

These sections correspond to homogeneous polynomials of degree n:

$$F(X_0, X_1) = \sum_{j=0}^n a_j X_0^{n-j} X_1^j, \qquad f_0 = F(1,t), \quad f_1 = F(u,1).$$

The homogeneous polynomial itself is not a scalar-valued function on projective points. Its chart expressions define a section of a line bundle, and the transition factor is what makes them compatible.

Why the sign of the transition matters

The tautological line over $[X_0 : X_1]$ is the line it spans in k^2. Local frames for this line bundle are $v_0 = (1, t)$ and $v_1 = (u, 1)$. On the overlap,

$$v_1 = t^{-1} v_0.$$

Thus the tautological bundle is $\mathscr{O}(-1)$ under our convention. Its dual has transition t and is $\mathscr{O}(1)$. This provides a concrete check on the signs in both the section formula and the earlier exceptional-normal-bundle calculation.

26.5 Čech cohomology on the projective line

Let $\mathfrak{U} = \{U_0, \dots, U_m\}$ be a finite ordered open cover and $\mathscr{F}$ a sheaf of abelian groups. Its *Čech cochain complex* has terms

$$C^q(\mathfrak{U}, \mathscr{F}) = \prod_{i_0 < \cdots < i_q} \mathscr{F}(U_{i_0} \cap \cdots \cap U_{i_q}), \qquad q \geq 0.$$

The differential is the alternating sum of restrictions:

$$(dc)_{i_0 \dots i_{q+1}} = \sum_{j=0}^{q+1} (-1)^j c_{i_0 \dots \widehat{i_j} \dots i_{q+1}} |_{U_{i_0} \cap \cdots \cap U_{i_{q+1}}}.$$

The hat means that the indicated index is omitted. In d^2, each double omission occurs twice with opposite signs, so $d^2 = 0$. Define

$$\check{H}^q(\mathfrak{U}, \mathscr{F}) = \ker d^q / \operatorname{im} d^{q-1},$$

where the image in degree zero is zero. Thus $\check{H}^0$ is the group of compatible local sections and equals $\Gamma(X, \mathscr{F})$ by the sheaf axiom. With two open sets, there are only terms C^0 and C^1: degree-one cohomology is the quotient of overlap sections by differences of sections from the two opens.

For the two-open cover $\mathfrak{U} = \{U_0, U_1\}$, the Čech cochain complex of $\mathscr{O}(n)$ is

$$C^0 = k[t] e_0 \oplus k[t^{-1}] e_1, \qquad C^1 = k[t, t^{-1}] e_0, \qquad C^i = 0 \ (i \geq 2).$$

Using the convention that the differential is "second restriction minus first restriction," we have

$$d(p(t) e_0, q(t^{-1}) e_1) = \big(t^n q(t^{-1}) - p(t)\big) e_0.$$

Its kernel is exactly the matching condition for global sections already computed. Its cokernel is

$$\check{H}^1(\mathfrak{U}, \mathscr{O}(n)) \cong \frac{k[t, t^{-1}]}{k[t] + t^n k[t^{-1}]}. \tag{26.1}$$

The monomial decompositions are

$$k[t] = \bigoplus_{j \geq 0} k t^j, \qquad t^n k[t^{-1}] = \bigoplus_{j \leq n} k t^j.$$

Thus

$$\check{H}^1(\mathfrak{U}, \mathscr{O}(n)) = \bigoplus_{n < j < 0} k\overline{t^j} = \begin{cases} 0, & n \geq -1, \\ \langle \overline{t^{n+1}}, \dots, \overline{t^{-1}} \rangle_k, & n \leq -2. \end{cases}$$

These classes are linearly independent because the denominator contains no monomial with $n < j < 0$. The nonzero case has dimension $-n-1$.

26.5.1 Sheaf cohomology and the comparison theorem

The groups just computed are defined using one cover. To state the cover-independent version, we define *sheaf cohomology*. A sheaf of abelian groups $\mathscr{I}$ is *injective* if every morphism $\mathscr{A} \to \mathscr{I}$ extends across any injective morphism $\mathscr{A} \hookrightarrow \mathscr{B}$. An *injective resolution* of $\mathscr{F}$ is an exact sequence

$$0 \longrightarrow \mathscr{F} \longrightarrow \mathscr{I}^0 \longrightarrow \mathscr{I}^1 \longrightarrow \mathscr{I}^2 \longrightarrow \cdots$$

with each $\mathscr{I}^q$ injective. Such resolutions exist. Applying global sections gives a cochain complex, and the definition is

$$H^q(X, \mathscr{F}) = H^q\big(\Gamma(X, \mathscr{I}^\bullet)\big).$$

Existence of resolutions and independence of this group from the chosen resolution are general sheaf-theoretic results that we use without proof. In degree zero, exactness gives $H^0(X, \mathscr{F}) = \Gamma(X, \mathscr{F})$. In higher degrees the groups detect the failure of global sections to preserve exactness. No injective resolution needs to be constructed for the calculation below.

We use two comparison results, also without proof; see [2, Chapter III, §§4–5].

Theorem 26.7 (Affine vanishing and affine-cover comparison)**.** *If $X = \operatorname{Spec} A$ and $\mathscr{F} = \widetilde{M}$ is quasi-coherent, then $H^q(X, \mathscr{F}) = 0$ for $q > 0$. More generally, if a scheme has a finite affine open cover $\mathfrak{U}$ such that every nonempty finite intersection of members is affine, then for every quasi-coherent sheaf $\mathscr{F}$,*

$$\check{H}^q(\mathfrak{U}, \mathscr{F}) \cong H^q(X, \mathscr{F}) \qquad (q \geq 0).$$

The hypotheses can be checked directly here: $U_0 = \operatorname{Spec} k[t]$, $U_1 = \operatorname{Spec} k[s]$, and $U_0 \cap U_1 = \operatorname{Spec} k[t, t^{-1}]$ are affine. The sheaf $\mathscr{O}(n)$ is locally free of rank one and hence quasi-coherent. The comparison theorem therefore identifies our computed quotient with $H^1(\mathbb{P}^1, \mathscr{O}(n))$ and yields

$$\dim_k H^0(\mathbb{P}^1, \mathscr{O}(n)) = \max\{n+1, 0\}, \qquad \dim_k H^1(\mathbb{P}^1, \mathscr{O}(n)) = \max\{-n-1, 0\}.$$

For a smooth projective geometrically integral curve C, the general comparison between arithmetic genus and sheaf cohomology is

$$g(C) = \dim_k H^1(C, \mathcal{O}_C).$$

This is another standard theorem, stated without proof; see [2, Chapter III, §5 and Chapter IV, §1]. The calculation above verifies it for $C = \mathbb{P}^1$, where $H^1(\mathbb{P}^1, \mathcal{O}) = 0$ and the genus is zero. For the elliptic curves of Lecture 23, it gives a one-dimensional group. The definition of genus used there did not require this later cohomological description.

The quotient records which discrepancies on the overlap can be removed by changing sections on the two charts. A nonzero class is an overlap discrepancy which cannot be removed. This calculation brings together the course's main ideas: localization supplies the overlap ring, modules describe the chartwise sections, transition functions define the geometry, and the cohomology of a short complex measures the failure of global gluing.

A full computation of the Cech quotient

With the frame e_0 on the overlap, the two-chart complex for $\mathcal{O}(n)$ is

$$k[t] \oplus k[t^{-1}] \xrightarrow{d} k[t, t^{-1}], \qquad d(f, g) = t^n g(t^{-1}) - f(t).$$

Its degree-one cohomology is the vector-space quotient

$$\frac{k[t, t^{-1}]}{k[t] + t^n k[t^{-1}]}.$$

The first summand of the denominator contains exactly the monomials with exponents at least zero. The second contains those with exponents at most n. If $n \geq -1$, every integer exponent lies in one of these two ranges, and the quotient is zero. If $n \leq -2$, the missing exponents are

$$n+1, n+2, \ldots, -1.$$

Their monomials span the quotient. They are independent because a sum of monomials with these exponents cannot be a sum of monomials from the two disjoint allowed ranges. Thus

$$h^1(\mathbb{P}^1, \mathcal{O}(n)) = \max\{-n-1, 0\}.$$

Here the identification of Cech cohomology with sheaf cohomology uses the comparison theorem stated earlier in this lecture; the quotient calculation itself is elementary linear algebra.

For $n = -2$, the class of $t^{-1}e_0$ is nonzero. There are no polynomials $f(t), g(t^{-1})$ with $t^{-1} = t^{-2}g(t^{-1}) - f(t)$, by the exponent argument. This gives an actual gluing discrepancy that cannot be removed by modifying sections on the two charts.

An exact sequence whose global sections lose surjectivity

Choose $p = [0:1] \in \mathbb{P}^1$. Multiplication by the section X_0 gives, for every integer n, an exact sequence

$$0 \to \mathscr{O}(n-1) \xrightarrow{X_0} \mathscr{O}(n) \longrightarrow k_p \to 0,$$

where the last identification includes a choice of basis of the one-dimensional fiber at p. On U_0, the multiplier is a unit in the local frames, so the cokernel vanishes. On U_1, it is multiplication by $u = X_0/X_1$, with cokernel $k[u]/(u)$ supported at p. This proves exactness chart by chart.

For $n = -1$, both $H^0(\mathscr{O}(-2))$ and $H^0(\mathscr{O}(-1))$ vanish, while $H^0(k_p) = k$. The long exact sequence therefore has the segment

$$0 \longrightarrow k \xrightarrow{\delta} H^1(\mathscr{O}(-2)) \longrightarrow H^1(\mathscr{O}(-1)) = 0.$$

Hence δ is an isomorphism. A value at p lifts locally to $\mathscr{O}(-1)$, but a nonzero value cannot lift to a global section. The failure is measured by precisely the one-dimensional Cech quotient above. This connects the local definition of exactness, the long exact sequence of Lecture 14, and an explicit projective calculation.

Finally the formulas give

$$\chi(\mathscr{O}(n)) = h^0(\mathscr{O}(n)) - h^1(\mathscr{O}(n)) = n+1$$

for every integer n. The same expression covers positive and negative twists because the contribution shifts from degree zero to degree one.

26.6 Exercises

Exercise 26.1 (Support and a nonfree stalk)**.** Let $A = k[x,y]$ and $M = A/(x)$. Determine the points where $\widetilde{M}$ has a nonzero stalk. Is $\widetilde{M}$ locally free on all of $\operatorname{Spec} A$?

Exercise 26.2 (A presentation becomes geometric)**.** Sheafify the exact sequence

$$0 \longrightarrow k[x] \xrightarrow{\cdot x} k[x] \longrightarrow k[x]/(x) \longrightarrow 0.$$

Describe the map at the origin and away from it.

Exercise 26.3 (Global sections and nontrivial line bundles)**.** Compute the global sections of $\mathscr{O}_{\mathbb{P}^1_k}(2)$ and $\mathscr{O}_{\mathbb{P}^1_k}(-1)$ on the projective line. Prove that $\mathscr{O}(-1)$ is not isomorphic to $\mathscr{O}$.

Exercise 26.4 (A visible gluing obstruction)**.** For the standard two-chart cover of $\mathbb{P}^1_k$, identify its first Čech cohomology group for $\mathscr{O}(n)$ with $k[t,t^{-1}]/(k[t] + t^n k[t^{-1}])$. In this quotient, compute the class of t^{-1} when $n = -2$, and the classes of t^{-1}, t^{-2}, t^{-3} when $n = -3$.

Exercise 26.5 (Sections of line bundles on the projective line)**.** Cover $\mathbb{P}^1$ by U_0 with coordinate $t = X_1/X_0$ and U_1 with coordinate $s = X_0/X_1 = t^{-1}$. For $\mathscr{O}(n)$ use frames e_0, e_1 with $e_1 = t^n e_0$. Compute $H^0(\mathbb{P}^1, \mathscr{O}(n))$ and the first Čech cohomology for this cover.

Exercise 26.6 (A coherent sheaf that is not locally free)**.** On $X = \mathbb{A}^2_k$, let $\mathscr{F} = \widetilde{(x,y)}$. Show that it is coherent and that it is a line bundle away from the origin, but is not a vector bundle on all of X.

Appendix A
Notation and a compact algebra–geometry dictionary

Notation	Meaning	Lecture
$A^\times$	Group of invertible elements of A.	1–2
$(f_1,\dots,f_r)$	Ideal of finite sums $\sum a_i f_i$; the ambient ring must be specified.	1–2
$\sqrt{I}$	Elements with some positive power in I.	2
$J(A)$	Intersection of all maximal ideals; the Jacobson radical.	2
M/N	Module obtained by imposing that every element of N is zero.	3
$\mathrm{Hom}_A(M,N)$	A-linear maps from M to N.	3–6
$M \otimes_A N$	Module representing A-balanced bilinear maps.	5
$S^{-1}A$, $S^{-1}M$	Ring and module in which denominators from S are allowed.	7
$A_\mathfrak{p}$, $M_\mathfrak{p}$	Localization with $S = A \setminus \mathfrak{p}$.	7–8
$\kappa(\mathfrak{p})$	Residue field $A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$.	8, 18
$\dim A$	Supremum of lengths of strict chains of prime ideals.	12
$\mathfrak{m}/\mathfrak{m}^2$	Cotangent space of a local ring over its residue field.	8, 12
$H^i(C^\bullet)$	$\ker d^i / \operatorname{im} d^{i-1}$: cycles modulo boundaries.	13–14
$V(I)$	Common zeros, or primes containing I, according to the stated ambient space.	16–17
$D(f)$	Basic open where f does not vanish; on $\mathrm{Spec}\,A$, primes not containing f.	16
$\mathrm{Spec}\,A$	Prime ideals with the Zariski topology; later equipped with a structure sheaf.	16, 25
$k[X]$, $A(X)$	Coordinate ring of an affine algebraic set (both notations are common).	17–18
$k(X)$	Fraction field of the coordinate ring of an irreducible affine variety; rational functions.	18, 21
$\mathbb{P}^n$	Nonzero $(n+1)$-tuples modulo common nonzero scaling, with projective geometry.	19
$P_M(n)$	Polynomial agreeing with the graded Hilbert function for sufficiently large n.	11, 20
$\widetilde{M}$	Sheaf associated to an A-module on $\mathrm{Spec}\,A$.	25–26

Notation	Meaning	Lecture
$\mathscr{F}_p$	Stalk: germs of local sections near p.	25
$\mathscr{O}_X$	Sheaf of regular functions on X.	25
$\mathscr{O}_{\mathbb{P}^1}(n)$	Line bundle with transition $e_1 = t^n e_0$ in the specified charts.	26

The algebra–geometry correspondence

Geometric statement	Algebraic statement
A closed subset of an affine algebraic set over an algebraically closed field	A radical ideal of its coordinate ring, with inclusions reversed.
An irreducible affine algebraic set	Its coordinate ring is a domain.
A point a in an affine algebraic set over an algebraically closed field	A maximal ideal of functions vanishing at a.
A regular map $X \to Y$ of affine varieties	A k-algebra map $k[Y] \to k[X]$, with the arrow reversed.
Concentrating near a point a	Localizing at its maximal ideal.
Restricting to the principal open $D(f)$	Inverting f.
Normalizing an integral affine curve	Taking integral closure in its function field.
A quasi-coherent sheaf on an affine scheme	An associated module; on a Noetherian affine scheme, coherence corresponds to finite generation.
A vector bundle of rank r	A sheaf locally isomorphic to $\mathscr{O}_X^{\oplus r}$.

The hypotheses are part of the dictionary. Over $\mathbb{R}$, the maximal ideal $(x^2+1) \subset \mathbb{R}[x]$ is not evaluation at a real number. On a nonaffine variety, a sheaf is not in general determined by its global sections. A bijection of point sets need not be an isomorphism of varieties, as the cusp illustrates.

Appendix B
Atiyah–Macdonald: source-result guide

This guide follows the numbering of the supplied edition of *Introduction to Commutative Algebra*. It covers the 155 numbered main-text items within the requested scope: propositions, theorems, lemmas, and corollaries, together with the six numbered in-text exercises and the starred Proposition 2.12*. Primary decomposition and completion are excluded. The remaining exercise sets are selected rather than exhaustive. Statements and proofs are written in original prose; the source numbering lets the reader check the exact correspondence. The six source in-text exercises are developed here as proved propositions. Most selected end-of-chapter exercises appear at lecture ends, with solutions in the companion volume; a few foundational exercises are developed as proved results in the main text.

Scope and exclusions

Chapter 4 is omitted. Also omitted are the primary-decomposition results 7.11–7.13, 7.17, and 9.1; and the completion results 10.1–10.5, 10.12–10.16, 10.22–10.27, 11.19, and 11.24. From Chapter 10 we retain 10.6–10.11 and 10.17–10.21: these concern graded rings, filtrations, Artin–Rees, and intersections of ideals. The elementary associated-graded finiteness argument from 10.22(i) and (iii) is proved separately; the complete completion-dependent statement 10.22 is excluded. Primary ideals in 7.16 and 9.3 are treated directly, without primary decomposition.

Results and proofs

Each page reference points to the statement in these notes. Its proof follows there or is explicitly linked to an earlier complete argument. Lecture numbers in these notes and chapter numbers in the source are different numbering systems.

AM	Kind	Subject	Lecture	Page
1.1	Proposition	Ideals in a quotient	2	24
1.2	Proposition	Three characterizations of a field	2	25
1.3	Theorem	Existence of maximal ideals	2	27
1.4	Corollary	Proper ideals lie in maximal ideals	2	27
1.5	Corollary	Nonunits lie in maximal ideals	2	27
1.6	Proposition	Two criteria for a local ring	2	27
1.7	Proposition	Nilradical and reduced quotient	2	28
1.8	Proposition	Nilradical as intersection of primes	2	28
1.9	Proposition	Jacobson radical unit criterion	2	29
1.10	Proposition	Chinese remainder theorem, all three parts	2	30
1.11	Proposition	Prime avoidance and finite intersections	2	31
1.12	Exercise	Five ideal-quotient identities	2	32
1.13	Exercise	Six radical identities	2	33
1.14	Proposition	Radical as intersection of containing primes	2	34
1.15	Proposition	Zero divisors as union of radical annihilators	2	34
1.16	Proposition	Comaximal radicals imply comaximal ideals	2	35
1.17	Proposition	Extension and contraction, all three parts	2	35
1.18	Exercise	Extension and contraction operations and closure	2	36
2.1	Proposition	Two quotient identities	3	49
2.2	Exercise (in text)	Annihilator identities	3	49
2.3	Proposition	Generators give a quotient of a free module	3	45
2.4	Proposition	The determinant trick for an endomorphism	8	95
2.5	Corollary	The determinant argument for generators	8	95
2.6	Proposition	Nakayama's lemma	8	96
2.7	Corollary	Nakayama for a quotient	8	96
2.8	Proposition	Lifting a basis of the residue module	8	97
2.9	Proposition	Both exactness tests	4	57
2.10	Proposition	Snake lemma	4	58
2.11	Proposition	Additive functions on an exact sequence	4	60
2.12	Proposition	Existence and uniqueness of tensor products	5	63
2.12*	Proposition	Several tensor factors	5	69
2.13	Corollary	A zero tensor has a finite witness	5	68
2.14	Proposition	Canonical tensor identities, all parts	5	69
2.15	Exercise (in text)	Mixed associativity	6	79
2.16	Proposition	Restriction and finite generation	6	78
2.17	Proposition	Extension and finite generation	6	78
2.18	Proposition	Tensor products preserve cokernels	5	66
2.19	Proposition	Equivalent flatness criteria	5	70
2.20	Exercise (in text)	Flatness after base change	6	79

AM	Kind	Subject	Lecture	Page
3.1	Proposition	Universal property	7	83
3.2	Corollary	A recognition criterion	7	87
3.3	Proposition	Exactness of localization	7	84
3.4	Corollary	Sums, intersections, and quotients	7	87
3.5	Proposition	Localization is extension of scalars	7	88
3.6	Corollary	Flatness of a ring of fractions	7	88
3.7	Proposition	Tensor products commute with localization	7	88
3.8	Proposition	Vanishing is local	7	89
3.9	Proposition	Injectivity and surjectivity are local	7	89
3.10	Proposition	Flatness is local	7	89
3.11	Proposition	Extended and contracted ideals, all parts	7	90
3.12	Corollary	Nilradicals localize	7	90
3.13	Corollary	Primes in a local ring of fractions	7	90
3.14	Proposition	Annihilators localize for finite modules	7	91
3.15	Corollary	Colon ideals localize under a finite-generation hypothesis	7	91
3.16	Proposition	When a prime is the contraction of a prime	7	91
5.1	Proposition	The four finite-module tests	10	124
5.2	Corollary	Finitely many integral generators	10	125
5.3	Corollary	Integral elements form a ring	10	125
5.4	Corollary	Transitivity	10	125
5.5	Corollary	Closing twice adds nothing	10	125
5.6	Proposition	Quotients and localization	10	126
5.7	Proposition	Fields in an integral extension	10	130
5.8	Corollary	Maximal ideals correspond under contraction	10	130
5.9	Corollary	Incomparability	10	130
5.10	Theorem	Lying over	10	130
5.11	Theorem	Going up	10	130
5.12	Proposition	Integral closure commutes with localization	10	131
5.13	Proposition	Normality is local	10	131
5.14	Lemma	The radical description	10	132
5.15	Proposition	Minimal-polynomial coefficients	10	132
5.16	Theorem	Going down	10	133
5.17	Proposition	Bounding an integral closure by a lattice	10	135
5.18	Proposition	Three properties of valuation rings	10	137
5.19	Lemma	The kernel of a maximal extension	10	137
5.20	Lemma	One of two extensions has a proper extended ideal	10	137
5.21	Theorem	Maximal extensions are valuation rings	10	138
5.22	Corollary	Integral closure as an intersection	10	138
5.23	Proposition	Extending maps while avoiding a zero	10	138
5.24	Corollary	The field form of the Nullstellensatz	10	139

AM	Kind	Subject	Lecture	Page
6.1	Proposition	Chain and maximal conditions	9	102
6.2	Proposition	Finite generation of all submodules	9	103
6.3	Proposition	Exact-sequence test	9	103
6.4	Corollary	Finite direct sums	9	104
6.5	Proposition	Finite modules over finite-chain rings	9	104
6.6	Proposition	Quotient rings	9	104
6.7	Proposition	Length is well-defined	9	112
6.8	Proposition	Finite length and the two chain conditions	9	112
6.9	Proposition	Additivity of length	9	112
6.10	Proposition	Vector spaces	9	113
6.11	Corollary	A product of maximal ideals equal to zero	9	113
7.1	Proposition	Surjective images	9	105
7.2	Proposition	Module-finite extensions	9	105
7.3	Proposition	Localization	9	106
7.4	Corollary	Local rings of a Noetherian ring	9	106
7.5	Theorem	Hilbert's basis theorem	9	106
7.6	Corollary	Finitely many variables	9	107
7.7	Corollary	Finitely generated algebras	9	107
7.8	Proposition	Artin–Tate lemma	9	108
7.9	Proposition	Zariski's lemma	9	110
7.10	Corollary	Weak Nullstellensatz	9	111
7.14	Proposition	A uniform power of the radical	9	114
7.15	Corollary	Nilradical of a Noetherian ring	9	114
7.16	Corollary	Ideals with maximal radical	9	114
8.1	Proposition	Primes in an Artinian ring	9	115
8.2	Corollary	The two radicals coincide	9	115
8.3	Proposition	Only finitely many maximal ideals	9	116
8.4	Proposition	Nilradical of an Artinian ring	9	116
8.5	Theorem	Artinian rings and dimension zero	9	116
8.6	Proposition	Powers of the maximal ideal	9	117
8.7	Theorem	Structure of Artinian rings	9	117
8.8	Proposition	Principal ideals in an Artinian local ring	9	118
9.2	Proposition	Six tests for a DVR	10	141
9.3	Theorem	The local criterion for a Dedekind domain	10	143
9.4	Corollary	Unique prime factorization of ideals	10	143
9.5	Theorem	Rings of algebraic integers	10	144
9.6	Proposition	Invertibility is local	10	146
9.7	Proposition	The local inverse-ideal test	10	146
9.8	Theorem	The global inverse-ideal test	10	147
9.9	Corollary	The group of fractional ideals	10	147

AM	Kind	Subject	Lecture	Page
10.6	Lemma	Stable filtrations differ by a bounded shift	11	158
10.7	Proposition	Graded Noetherian rings	11	154
10.8	Lemma	Stability is finite generation	11	159
10.9	Proposition	Artin–Rees, stable-filtration form	11	159
10.10	Corollary	Artin–Rees, intersection formula	11	159
10.11	Theorem	Bounded difference of the submodule filtrations	11	160
10.17	Theorem	Krull's intersection theorem	11	160
10.18	Corollary	Intersection in a domain	11	160
10.19	Corollary	Intersection for a Jacobson-radical ideal	11	160
10.20	Corollary	Intersection in a local ring	11	161
10.21	Corollary	Intersection of primary neighborhoods	11	161
11.1	Theorem	Hilbert–Serre	11	155
11.2	Corollary	Eventual polynomiality	11	156
11.3	Proposition	Removing a homogeneous non-zero-divisor	11	156
11.4	Proposition	Hilbert–Samuel polynomial for a filtered module	12	164
11.5	Corollary	The characteristic polynomial	12	165
11.6	Proposition	The degree is independent of the ideal	12	165
11.7	Proposition	Equations bound growth	12	166
11.8	Proposition	A quotient lowers the degree	12	166
11.9	Corollary	Growth of a hypersurface quotient	12	166
11.10	Proposition	Growth bounds every prime chain	12	166
11.11	Corollary	Local dimension is finite	12	167
11.12	Corollary	Finite heights and descending prime chains	12	167
11.13	Proposition	Enough equations can always be chosen	12	168
11.14	Theorem	The local dimension theorem	12	168
11.15	Corollary	Dimension is bounded by first-order directions	12	169
11.16	Corollary	The generalized height theorem	12	169
11.17	Corollary	Krull's principal ideal theorem	12	169
11.18	Corollary	A non-zero-divisor cuts dimension by one	12	169
11.20	Proposition	Independence of parameters	12	171
11.21	Corollary	Parameters over a coefficient field	12	171
11.22	Theorem	Equivalent tests for regularity	12	172
11.23	Lemma	A domain criterion from the associated graded ring	12	173
11.25	Theorem	Local dimension equals transcendence degree	12	178
11.26	Lemma	Local dimension under an integral extension	12	177
11.27	Corollary	Global dimension agrees with every closed local dimension	12	178

Selected exercises from the book

The source chapter is included because exercise numbering restarts in each chapter. If an exercise is used in part, the statement in the lecture specifies the selected subparts. These are additional to the original examples and practice problems in the notes. A page reference locates either a lecture-end exercise statement or, when its result is needed later, the corresponding proved result in the exposition. Lecture-end exercise solutions use the same exercise numbers in the separate companion PDF. In particular, source Exercises 2.2 and 5.16 are developed as proved main-text results because the quotient tensor formula and Noether normalization are used later.

AM exercise	Topic	Page
1.1	Nilpotent plus unit; both assertions	40
1.2	Polynomial units, nilpotents, zero divisors, and primitive products; all parts (i)–(iv)	40
1.4	Jacobson radical equals nilradical in a polynomial ring	40
1.11	Boolean rings; all parts (i)–(iii)	40
1.12	A local ring has no nontrivial idempotent	41
1.15	Zariski closed sets; all parts (i)–(iv)	41
1.17	Basic open sets; all parts (i)–(vii)	41
1.22	Finite product spectrum and disconnectedness/product/idempotent equivalences	41
2.1	Integer and polynomial computations	71
2.2	Quotient formula	65
2.3	Nonzero finite modules over a local ring have nonzero tensor product	101
2.4	Direct sums and flatness	71
2.6	Polynomials with module coefficients	80
2.8	Two permanence properties for flat modules	80
2.9	Finite generation in an extension	61
2.10	Surjectivity after reduction	101
3.1	A uniform annihilating denominator	101
3.2	Moving an ideal into the Jacobson radical	101
3.3	Two successive choices of denominators	92
3.5	Reducedness is local, but being a domain is not	92
3.12	Torsion and fractions	92
3.13	Torsion-free is a local property	92
3.19	All the basic support identities	101
5.1	Integral maps are closed	149
5.2	Extending a point through an integral extension	149
5.3	Integrality under base change	149
5.4	A localization trap	149
5.12	Finite symmetries give integral extensions	149
5.16	Noether normalization	139
6.1	Surjective and injective endomorphisms	120

AM exercise	Topic	Page
6.2	A restricted maximal condition suffices	120
6.3	Two quotient tests	120
6.4	The annihilator ring	120
7.1	Cohen's finiteness criterion	120
7.5	Invariant rings of finite groups	120
7.6	Fields finitely generated as rings	120
8.2	A discrete spectrum	120
8.3	Finite type plus dimension zero	120
8.4	Discrete fibers need not make a map finite	120
9.2	Content and Gauss's lemma	149
9.3	When a valuation ring is Noetherian	149
9.4	A principal maximal ideal and separated powers	149
9.5	Flat modules over a Dedekind domain	150
9.7	Two generators suffice	150
9.8	Distributivity of ideal arithmetic	150
9.9	Compatible congruences without coprimality	150
11.1	Hypersurface singularities and the derivative criterion	179
11.3	Dimension over an arbitrary ground field	179
11.6	Polynomial dimension bounds for arbitrary rings	179
11.7	Polynomial dimension for Noetherian rings	179

References

1. M. F. Atiyah and I. G. Macdonald, *Introduction to Commutative Algebra*, Addison–Wesley, 1969. Supplied reference PDF; chapter citations refer to the printed book, not to the scanned spread numbers.
2. R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Mathematics 52, Springer, 1977. Supplied reference PDF. Roman numerals in references indicate chapters; for example II.5 means Chapter II, Section 5.
3. Si Li, *Geometry and Symmetry* (in Chinese), lecture notes for the Fall 2018 course at Tsinghua University. Supplied file `geometry_symmetry.pdf`; especially Sections 1.1–1.4 and 2.1.
4. The Chinese University of Hong Kong, Shenzhen, *MATH3005: Abstract Algebra II*, supplied course outline, version V001. The teaching sequence in these notes follows its indicative Weeks 1–15 plan, with the requested opening algebra review.
5. J. S. Milne, *Elliptic Curves*, 2006. Supplementary arithmetic reading, especially Chapter IV on elliptic curves over the rational numbers. Author's text: https://www.jmilne.org/math/Books/ectext6.pdf.