0
Hello!

This chapter contains a preliminary discussion of the aims and philosophy of the book, and is not logically part of the course. Some of the material may be harder to follow here than when it is treated more formally later in the book, so if you get stuck on something, don't worry too much, just skip on to the next item.

0.1 Where we're going

The purpose of this course is to build one of the bridges between algebra and geometry. Not the Erlangen program (linking geometries via transformation groups with abstract group theory) but a quite different bridge linking rings A and geometric objects X; the basic idea is that it is often possible to view a ring A as a certain ring of functions on a space X, to recover X as the set of maximal or prime ideals of A, and to derive pleasure and profit from the two-way traffic between the different worlds on each side.

Algebra here means rings, always commutative with a 1, and usually closely related to a polynomial ring $k[x_1, \dots, x_n]$ or $\mathbb{Z}[x_1, \dots, x_n]$ over a field k or the integers $\mathbb{Z}$, or a ring obtained from one of these by taking a quotient by an ideal, a ring of fractions, a power series completion, and so on; also their ideals and modules. In this book, A usually stands for a ring and k for a field, and I sometimes use these notations without comment. We're interested in questions such as zerodivisors (that is, $0 \neq x, y \in A$ such that $xy = 0 \in A$), factorisation (that is, writing $a \in A$ as a product $a = b \prod p_i^{n_i}$ with b invertible and p_i prime elements), similar questions for ideals, prime ideals, extension rings $A \subset B$, etc.

The study of rings of this type includes most of algebraic number theory and a large fraction of algebraic geometry. The methods for

studying them, by and large, are either simple algebraic arguments, or depend on the link with geometry which I want to introduce here. Thus, for example, rings have a *dimension theory*, in which $\dim k[x_1, \dots, x_n] = n$ and $\dim \mathbb{Z}[x_1, \dots, x_n] = n + 1$ (yes, $n + 1$ is right!), and already the language suggests a cocktail of two different subjects. The same holds for *local ring*, an idea at the very heart of commutative algebra.

0.2 Some definitions

Before describing briefly the geometric side and some aspects of the bridge, I recall a few very elementary algebraic topics and introduce some definitions, which I hope are mostly already familiar.

Let A be a ring, commutative with a 1. *Zerodivisors* of A are nonzero elements $x, y \in A$ such that $xy = 0$. If A has no zerodivisors and $A \neq 0$, it is an *integral domain*; note that $0 \neq 1$ is part of the definition of integral domain. An integral domain is contained in a unique field K such that every element of K is a fraction a/b with $a, b \in A$ and $b \neq 0$; this is the *field of fractions* of A, sometimes written $K = \operatorname{Frac} A$, and I assume you understand its construction. An element $x \in A$ is *invertible* or a *unit* of A if it has an inverse in A, that is, there exists $y \in A$ such that $xy = 1$.

An element $x \in A$ is *nilpotent* if $x^n = 0$ for some n. Prove for yourself that x nilpotent implies that $1 - x$ is invertible in A. [Hint: write out $(1 - x)^{-1}$ as a power series.] Prove also that x and y nilpotent implies that $ax + by$ is nilpotent for all $a, b \in A$, so that the set of nilpotent elements of A is an ideal, the *nilradical* $\operatorname{nilrad} A$. [Hint: use the binomial theorem.] An element $x \in A$ is *idempotent* if $x^2 = x$. Obviously if x is idempotent then so is $x' = 1 - x$, and then $x + x' = 1$ and $xx' = 0$ (please check all this for yourself), so that x and x' are *complementary orthogonal idempotents*; now by writing $a = ax + ax'$ for any $a \in A$, you see that A is a direct sum of rings $A = A_1 \oplus A_2$, where $A_1 = Ax$ and $A_2 = Ax'$.

0.3 The elementary theory of factorisation

Suppose that A is an integral domain. A nonzero element $x \in A$ is *irreducible* if x itself is not invertible, and $x = yz$ with $y, z \in A$ implies that either y or z is invertible. $x \in A$ is a *prime element* if it is not a unit, and $x \mid yz$ implies either $x \mid y$ or $x \mid z$. It is trivial to see that prime implies irreducible, but the other way round is false in general.

A is a *UFD* (*unique factorisation domain*) if (i) every element x factors as a product of finitely many irreducibles $x = \prod x_i$ with x_i irreducible, and (ii) irreducible implies prime.

Proposition *In a UFD A, the expression of $x = b\prod p_i^{n_i}$ as a product of irreducibles with $p_i \nmid p_j$ is unique (up to invertible elements).*

I assume you know this. Otherwise, see any textbook on algebra, for example, [C], [H & H], Chapter 4 or [W]. In the following sections, I need to assume that the polynomial ring $k[x_1, \dots, x_n]$ is a UFD; the proof of this is discussed in Exs. 0.8–9 below.

0.4 A first view of the bridge

For simplicity, and to be able to describe in a few intuitive words a representative case of the geometric side, suppose that k is an algebraically closed field, for example $k = \mathbb{C}$. Then the polynomial ring $k[x_1, \dots, x_n]$ is a ring of functions on k^n, because a polynomial $g \in k[x_1, \dots, x_n]$ is a function $g = g(x_1, \dots, x_n)$ of $x_1, \dots, x_n$. Moreover, evaluating a polynomial at a point $P = (a_1, \dots, a_n) \in k^n$ determines a homomorphism $k[x_1, \dots, x_n] \to k$ defined by $g \mapsto g(P)$, whose kernel is the maximal ideal $m_P = (x_1 - a_1, \dots, x_n - a_n)$ (see Ex. 1.15). This is the correspondence between a ring A and a space X in ideal form: $A = k[x_1, \dots, x_n]$ is the ring of polynomial functions on $X = k^n$, and the points of X correspond to maximal ideals of A.

0.5 The geometric side – the case of a hypersurface

Suppose that $0 \neq F \in k[x_1, \dots, x_n]$; then the locus

$$X = V(F) = \{P = (a_1, \dots, a_n) \mid F(P) = 0\} \subset k^n$$

is a hypersurface. It is $(n-1)$-dimensional because you can (almost always) use the equation $F = 0$ to solve for a_1 in terms of $a_2, \dots, a_n$.

Now consider the quotient ring $A = k[x_1, \dots, x_n]/(F)$, that is, the ring of residue classes modulo the ideal generated by F. Then an element $g \in A$ defines a k-valued function on X: indeed, if g is the class in A of a polynomial $\widetilde{g} \in k[x_1, \dots, x_n]$ then for $x \in X$ the value $g(x) = \widetilde{g}(x)$ does not depend on the choice of $\widetilde{g}$.

Now we can see (fairly light-weight) traffic passing over the bridge. First, to what extent can A be viewed as a ring of functions on X?

(i) If F has no multiple factors, say $F = \prod f_i$ with $f_i \nmid f_j$, then it can be shown that F generates the ideal of all functions vanishing on X. (You can try the proof as an exercise after Chapter 5, see Ex. 5.5.) It follows from this that an element $g \in A$ is uniquely determined by the corresponding function $g\colon X \to k$, so that A is contained in the ring of k-valued functions on X.

On the other hand, if F has a multiple factor, say $F = f^k g$ with $k \geq 2$ then also $fg = 0$ everywhere on X, and hence F does not generate the ideal of all functions vanishing on X. At the same time, A has nonzero nilpotent elements (because $x = \operatorname{im} fg \in A$ satisfies $x^k = 0$). In this case, it is not reasonable to try to view the nilpotent element x as a function on X, because it is zero everywhere on X. Thus

F has a multiple factor

$\iff$ more functions vanish on X than (F)

$\iff$ A has nilpotent elements

$\iff$ A has nonzero elements that are 0 as functions on X.

(ii) If F has a factorisation $F = f_1 f_2$, where f_1, f_2 are polynomials with no common factors, then A has zerodivisors (because $x_1 = \operatorname{im} f_1$, $x_2 = \operatorname{im} f_2$ satisfy $x_1 \neq 0, x_2 \neq 0$ but $x_1 x_2 = 0$); this corresponds to a decomposition $X = X_1 \cup X_2$ of X as a union of two smaller hypersurfaces X_i given by $f_i = 0$ for $i = 1, 2$. Thus

A has zerodivisors (not nilpotents)

$\iff$ X is reducible: $X = X_1 \cup X_2$.

That is, something in algebra equals something in geometry.

(iii) I mentioned complementary orthogonal idempotents and direct sums of rings in 0.2; you can't get much more abstract algebraic than that. However, it is easy to see that A has nontrivial idempotents if and only if X is a disjoint union of two hypersurfaces, $X = X_1 \sqcup X_2$. If $k = \mathbb{C}$, this just means that $X \subset \mathbb{C}^n$ is a disconnected topological space; you can't get much more geometric than that. The ring of functions (say, continuous) on a disconnected space $X = X_1 \sqcup X_2$ is a direct sum of the rings of functions on X_1 and X_2.

(iv) We will see that there is a close relation between ideals $I \subset A$ and subvarieties of X; we can already see that if $I \subset A$ is an ideal then it defines a *subvariety* $V(I) \subset X$, the subset of $P \in X$ where

$f(P) = 0$ for all the functions $f \in I$. But this is quite a long story that I defer until later. For the time being, I state without proof the following result (a special case of the weak Nullstellensatz, see Theorem 4.10 and 5.1).

Proposition *Maximal ideals of A are in one-to-one correspondence with points $P \in X$. That is,*

$$P = (a_1, \dots, a_n) \in X \leftrightarrow m_P = (x_1 - a_1, \dots, x_n - a_n) \subset A.$$

To repeat my refrain, something in algebra (the maximal ideals of A) equals something in geometry (the points of X).

(v) Assume that $F \in k[x_1, \dots, x_n]$ is irreducible, so that A is an integral domain. When is $A = k[x_1, \dots, x_n]/(F)$ a UFD?

For example, if $F = xz - y^2 \in k[x, y, z]$ then $xz = y^2$ holds in the quotient ring A, whereas it is not hard to check that x, y, z are irreducible; therefore A is not a UFD. Now draw the picture of the locus $X : (xz = y^2)$, which is the ordinary quadric cone (see Figure 0.5). I will come back to this picture several times later in the book. Observe that X is a cone, and so contains lots of lines, for example, the lines $L_\lambda \subset X$ defined by $x = \lambda^2 z, y = \lambda z$; these are codimension 1 subvarieties of X.

I take $\lambda = 0$ to simplify the notation, so consider the line $L = L_0 \subset X$ defined by $x = y = 0$. The special feature of X is that the ideal $I_L \subset A$ of functions vanishing on L is not generated by one element. In fact $I_L = (x, y)$, but y also vanishes along a second line $y = z = 0$, whereas x vanishes along L with multiplicity 2. Geometrically, this corresponds to the fact that the plane $x = 0$ is everywhere tangent to X along L; or, to put it another way, at any point where $z \neq 0$, I have $x = y^2/z$. In this sense, L is not locally defined by one equation.

In other words, the geometric question of codimension 1 subvarieties and how to define them by equations is closely related to the algebraic question of unique factorisation in the ring A.

0.6 $\mathbb{Z}$ versus $k[X]$

The comparison between the ring of integers $\mathbb{Z}$ and the polynomial ring $k[X]$ in a single variable over a field k is one of the central points to be made at the outset in a commutative algebra course. From the algebraic

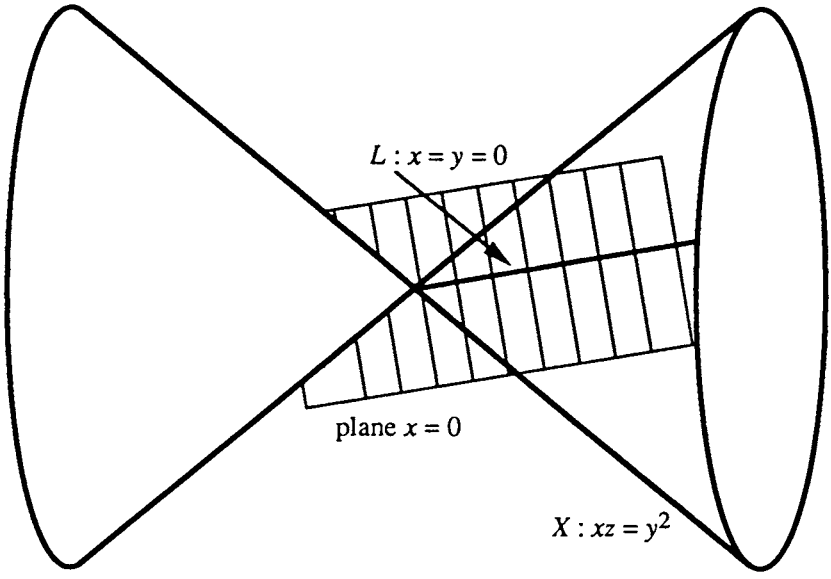

Figure 0.5. Quadric cone with a line

point of view, these two rings are very similar in many formal respects, and yet they are very different in substance. (Compare also Exs. 0.10–12 and the worked example 1.5.)

Points of similarity Recall that $\mathbb{Z}$ and $k[X]$ are both *Euclidean rings*, that is, integral domains satisfying "division with remainder" or the "Euclidean algorithm": for any a, b, there exists an expression $a = bq + r$ with r less than b. The ideal theory of the two rings proceeds in parallel from this fact: from division with remainder it follows easily that every ideal is *principal* (generated by one element), either (0) or (f) for an element f. I assume you know all this (see Exs. 0.1–9), including how you deduce the familiar $af + bg = h$ property of the highest common factor $h = \mathrm{hcf}(f, g)$, unique factorisation, etc.; if not, see, for example, [C] or [H & H], Chapter 4.

Points of difference Obviously $\mathbb{Z}$ and $k[X]$ are as different as chalk and cheese, but it is worthwhile trying to pin down the difference. I give two illustrations. First an algebraic statement: $k[X]$ contains a field, whereas $\mathbb{Z}$ doesn't. As you know, for any ring A, there is a unique

homomorphism $t\colon \mathbb{Z} \to A$: after taking $1 \mapsto 1_A$, the rest is forced. If A contains a field then t factors through the prime subfield, either $\mathbb{F}_p$ or $\mathbb{Q}$. If a ring A contains a field k as a subring, then A and any A-module are k-vector spaces. In the case $A = \mathbb{Z}$, there is obviously no way of embedding either $\mathbb{F}_p$ or $\mathbb{Q}$ into A. The same holds for $A = \mathbb{Z}/(p^2)$, with p a prime number: the additive group $\mathbb{Z}/(p^2)$ is not a vector space over any field; see Ex. 0.10. For this reason, one sometimes says that a ring containing a field k has *equal characteristic* $\operatorname{char} k$ (either 0 or p), whereas a ring like $\mathbb{Z}$ has *unequal characteristic*, in the sense that the ring itself has characteristic 0, but it has residue fields $\mathbb{F}_p$ of characteristic p.

Here is another difference: $k[X]$ contains *variables*. To put it algebraically, a typical maximal ideal of $k[X]$ is (X), and it makes sense to *differentiate* with respect to X; that is, there is a k-linear map

$$\frac{\mathrm{d}}{\mathrm{d}X}\colon k[X] \to k[X], \qquad \text{defined by } x^n \mapsto nx^{n-1} \text{ for all } n,$$

with the properties and applications that you know about. By contrast, the maximal ideals of $\mathbb{Z}$ are (2), (3), (5), etc., and

$$\frac{\mathrm{d}}{\mathrm{d}2}, \frac{\mathrm{d}}{\mathrm{d}3}, \frac{\mathrm{d}}{\mathrm{d}5}, \dots$$

are of course completely meaningless. There is no nonzero derivation of $\mathbb{Z}$ to anything.

To put it another way, multiplication by a natural number n is the additive operation $a \mapsto a + \cdots + a$ (with n summands), and therefore this operation, and the ideal (n) generated by n, are already determined by the additive structure.

0.7 Examples

Recall that we intend to study extension rings of $\mathbb{Z}$ and of k, and that the distinctions in 0.6 will carry over to these. I continue the theme of 0.6 with two slightly more substantial examples from algebraic geometry and algebraic number theory illustrating this and other points.

Example 1 Suppose that k is an algebraically closed field of characteristic $\neq 2$, and let A be the ring $A = k[X,Y]/(Y^2 - X^3)$. By the correspondence of 0.5, A is a ring of functions on the plane curve $C \subset k^2$ given by $Y^2 = X^3$. See Figure 0.7(a).

Now A can be viewed as the extension ring $k[X][\sqrt{X^3}]$ obtained by adjoining the square root of X^3 to $k[X]$. If $(X - \alpha)$ is a maximal ideal

of $k[X]$ with $\alpha \in k$, it is contained in maximal ideals $(X-\alpha, Y-\beta)$, corresponding to the square roots $\beta = \pm\sqrt{\alpha^3}$; there are obviously two of these if $\alpha \neq 0$, and one otherwise. These ideals require two generators. Moreover, it is easy to see that the elements X and Y are irreducible in A, but not prime: indeed, $Y^2 = X^3$ in A, so that $X \mid Y^2$, but $X \nmid Y$.

At this point, observe that I am doing something fairly silly by taking the square root of X^3. It is clearly much more sensible to take the square root of X instead. Let $A' = k[t]$ where $t = Y/X = \sqrt{X}$; this is a slightly bigger ring that A. Then in it, $X = t^2$ and $Y = t^3 \in A' = k[t]$, so that A' is just a polynomial ring, so of course it is a UFD, and every ideal is principal.

Example 2 Now consider $B = \mathbb{Z}[\sqrt{-3}\,]$, the extension ring obtained by adjoining $\sqrt{-3}$ to $\mathbb{Z}$. What are its maximal ideals? If P is a nonzero prime ideal of B then $P \cap \mathbb{Z}$ is a nonzero prime ideal of $\mathbb{Z}$, so of the form (p); we say that P *lies over* p. I check first that every prime number $p \neq 2, 3$ either splits as a product $p = f_+ f_-$ of two prime elements of B, or remains a prime element of B, and, in particular, any prime ideal of B not lying over 2 is principal. Indeed, any $p \equiv 1 \bmod 6$ can be written as $p = 3a^2 + b^2$ with $a, b \in \mathbb{Z}$; this was proved in 1760 by Euler, and is worked out in Ex. 0.14. Thus p factors in B as a product of two irreducible elements $p = f_+ f_-$, with $f_\pm = b \pm a\sqrt{-3}$; for example,

$$7 = (2 + \sqrt{-3}\,)\,(2 - \sqrt{-3}\,).$$

It is easy to check that $B/(f_\pm) \cong \mathbb{F}_p$, so that $f_\pm$ are two prime elements of B lying over p. If $p \equiv 5 \bmod 6$ then $X^2 + 3$ is irreducible in $\mathbb{F}_p[X]$ by quadratic reciprocity (see Ex. 0.13), and $B/(p) \cong \mathbb{F}_p[X]/(X^2+3) \cong \mathbb{F}_{p^2}$ is a field, so that p is a prime element of B. It is also not hard to check that $(\sqrt{-3}\,) = 3\mathbb{Z} \oplus \mathbb{Z}\sqrt{-3} \subset B$, so that $B/(\sqrt{-3}\,) = \mathbb{F}_3$, and again $\sqrt{-3}$ is a prime element of B.

However, 2 is bad in B: you can easily prove it is irreducible, because

$$(a + b\sqrt{-3}\,)(a - b\sqrt{-3}\,) = a^2 + 3b^2 \neq 2 \qquad \text{for any } a, b \in \mathbb{Z},$$

but it is not a prime element since $2^2 = (1 + \sqrt{-3}\,)(1 - \sqrt{-3}\,)$. Thus the prime ideal over 2 in B is $(2, 1 + \sqrt{-3}\,)$, which needs 2 generators.

At this point, anyone who knows algebraic number theory will see that I am doing something fairly silly by taking the square root of -3. It is clearly much more sensible to take $\omega = (-1+\sqrt{-3}\,)/2$ instead; note that $\omega = \exp(2\pi i/3)$ is a primitive cube root of 1, satisfying $\omega^2 + \omega + 1 = 0$. Let $B' = \mathbb{Z}[\omega]$, a slightly bigger ring than B. The analysis of its prime

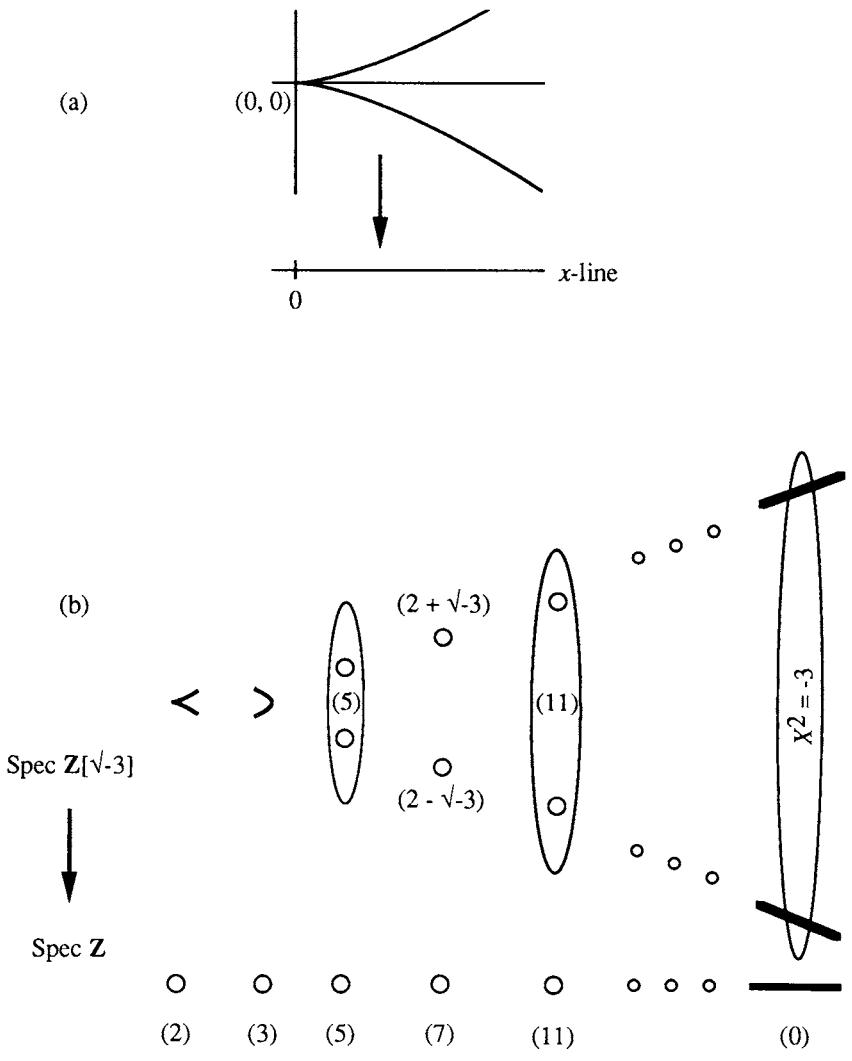

Figure 0.7. The cuspidal cubic and Spec $\mathbb{Z}[\sqrt{-3}]$

ideals is exactly as for B except above 2, which is a prime element of B', because $B'/(2) = \mathbb{F}_2[\omega] = \mathbb{F}_4$, a quadratic extension field of $\mathbb{F}_2$. In fact you can prove that B' is a UFD (see Ex. 0.16).

Figure 0.7(b) draws the prime ideals of $B = \mathbb{Z}[\sqrt{-3}]$ in schematic form. I draw two points in a bubble over the primes $p \equiv 5 \bmod 6$ to represent the single prime $p \in B$, because I have in mind the two conjugate points $X = \pm\sqrt{-3}$ of the X-line defined over $\mathbb{F}_{p^2}$.

0.8 Reasons for studying commutative algebra

Commutative algebra is the crossroads between algebraic number theory, algebraic geometry and abstract algebra. Although much of the material of this book develops techniques of algebra, it should be clear that my main interest is the applications of these ideas to geometry and number theory.

a. Algebraic number theory Galois theory studies field extensions, with motivation coming from the study of polynomials and their roots; thus, corresponding to a polynomial

$$f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_0 \in k[X]$$

with coefficients in a field k, one knows how to build a field extension $k \subset K$ over which f has a root, or splits into linear factors. It often happens that k contains a subring $A \subset k$ of interest, for example $\mathbb{Z} \subset \mathbb{Q}$, such that the coefficients of f are contained in A; we might then want to study a subring of K corresponding to A, for example the subring B generated by A and the roots of f. As a famous example, let $\varepsilon = \exp(2\pi i/n) = \sqrt[n]{1} \in \mathbb{C}$; over the ring $B = \mathbb{Z}[\varepsilon]$, a number of the form $x^n - z^n$ with x and z coprime factorises into n factors:

$$x^n - z^n = \prod_{i=0}^{n-1} (x - \varepsilon^i z).$$

Suppose we happened to know that $B = \mathbb{Z}[\varepsilon]$ is a UFD; then comparing two factorisations into primes coming from the left and right sides of $\prod_{i=0}^{n-1}(x - \varepsilon^i z) = y^n$ would obviously impose very strong restrictions on integer solutions of $x^n + y^n = z^n$, and in fact it is known that this can be used to prove that there are only the trivial solutions with x or y or $z = 0$; see, for example, [B & Sh], Chapter III, 1.1. Unfortunately, $\mathbb{Z}[\varepsilon]$