Differential Arithmetic

R.C. Churchill

Prepared for the
Kolchin Seminar on Differential Algebra
(KSDA)
Department of Mathematics, Graduate Center, CUNY
February, 2008

In this talk we indicate how elementary number-theoretic results depending on unique factorization can occasionally be established by "differentiating" integers. The presentation is intended to serve as an introduction to differential algebra, and is kept at a very elementary level: no background in that area is assumed, and unique factorization is thoroughly reviewed. For deep connections between number-theory and differential and difference algebra readers should consult [Bui, Hru, H-P, Scan] and references therein.

Contents

Introduction

Differential algebra is rooted in differential equations, but the techniques are purely algebraic. This latter fact can present difficulties for those who would like to learn the subject, but who formulate their mathematics in other ways.

In this talk we have chosen to fill some of the posssible algebraic gaps by grouping the necessary prerequisites around the topic of Diophantine equations, i.e., the search for integer solutions of equations with integer coefficients. Of course there may be no such solutions, e.g., there is certainly no integer solution of $2x+5=6$. Of interest for our purposes is the long acknowledged dictum that when one can establish existence the most incisive argument frequently involves some ring other than $\mathbb{Z}$. In this lecture we illustrate this assertion by means of the following three problems.

I. Show that $x^3+y^3=z^3$ has no solutions $x,y,z\in\mathbb{Z}$ with $xyz\neq 0$ and none of x,y and z divisible by 3. This is a special case of *Fermat's Last Theorem*: the assertion that for any integer $n\geq 3$ the equation $x^n+y^n=z^n$ has no integer solutions satisfying $xyz\neq 0$. The problem is easily reduced to the case in which n is a prime p, and two alternatives are then considered: CASE I - p does not divide xyz; CASE II - p divides xyz. Problem I. can therefore be described as Case I of Fermat's Last Theorem for $p=3$.

II. Find all solutions of $x^2+y^2=z^2$ with $x,y,z\in\mathbb{Z}$ and $xyz\neq 0$. Solutions are called *Pythagorean triples* (for hopefully obvious reasons).

III. Find necessary and sufficient conditions on a positive integer n to guarantee that n can be written as a "sum of two squares," i.e., expressed in the form $n=x^2+y^2$ with $x,y\in\mathbb{Z}$.

Ring-theoretic prerequisites which would normally be encountered in an undergraduate or first-year graduate algebra course are stated without proof, but always with at least one reference to a proof.

1. Elementary Ring Constructions

God made the integers; all else is the work of man.
Leopold Kronecker

We take "the integers" to mean "the ring $\mathbb{Z}$."

The "work of man" begins with the construction of auxiliary rings. It is worth recalling two such constructions explicitly[1], and in so doing we can avoid later digressions by replacing the usual ring of integers $\mathbb{Z}$ with an arbitrary commutative ring A with unity (i.e., multiplicative identity) 1.

Our presentation is somewhat tongue-in-cheek: in keeping with Kronecker's statement we proceed as if totally ignorant of the field of rational numbers. Indeed, that field will be one of the rewards of our efforts. What we do assume familiar are the definitions of ring, ring homomorphism, the kernel of a ring homomorphism $f : A \to B$, i.e., the inverse image of the zero element $0 = 0_B \in B$, ring isomorphism, module[2],

[1] The residue class ring construction of (a), if not that of a ring of fractions in (b), is assumed familiar to readers. Our immediate purpose is more to establish notation and a particular viewpoint than to present new material.

[2] "Module" will always mean "left module" unless specifically stated to the contrary. Readers comfortable with modules should skip the rest of this footnote; others may find a review of the definition useful.

Let A be a ring, not necessarily commutative, not necessarily with unity, and let M be an abelian group, written additively. If A admits a unity, denote that unity by 1_A. M is a *left A-module* if there is a function $(a, m) \in A \times M \mapsto am \in M$ such that for all $a, b \in A$ and all $m, n \in M$ one has

$$\begin{array}{lllll} \text{(i)} & (a+b)m & = & am + bm, & \\ \text{(ii)} & a(m+n) & = & am + an, & \\ \text{(iii)} & (ab)m & = & a(bm), & \text{and} \\ \text{(iv)} & 1_A m & = & m & \text{if } A \text{ admits a unity,} \end{array}$$

When confusion might otherwise result we write am as $a \cdot m$. Examples: the even integers (with multiplication ignored) form a $\mathbb{Z}$-module (when multiplication as usual is understood); when A is a field any vector space over A is an A-module.

Some authors would define a left A-module using only (i)-(iii), even if A admits a unity; A-modules satisfying (iv) would then be called *unitary left A-modules*. For an example of a non-unitary left A-module let M be any abelian group and define $a \cdot m := 0 \ (\in M)$ for all $(a, m) \in A \times M$.

The definition of a right A-module is completely analogous: replace the function $(a, m) \mapsto am$ by a function $(m, a) \mapsto ma$ and reverse the order in (i)-(iii), e.g., replace (ii) by $(m+n)a = ma + na$.

A left A-module can also be a right A-module, e.g., any ring B containing A as a subring can be considered as such by ignoring multiplication in B and defining $a \cdot b := ab$ and $b \cdot a := ba$. As one can see from simple matrix examples, $ab = ba$ will not automatically hold.

module homomorphism[3], algebra[4], algebra homomorphism[5], the kernel of an alge-

[3]Let A be a ring and let M and N be A-modules. A(n additive) group homomorphism $f : M \to N$ is an *A-module homomorphism* if $f(am) = af(m)$ for all $(a, m) \in A \times M$. Example: When A is a field any transformation between vector spaces over A is an A-module homomorphism.

[4]As was the case with modules, readers comfortable with the notion of an algebra should skip this footnote.

Let A be a commutative ring. By *A-algebra* we simply mean an A-module M together with an A-bilinear mapping $\mu : M \times M \to M$, i.e., a function μ such that for each $n \in M$ the mappings $m \in M \mapsto \mu(m, n) \in M$ and $m \in M \mapsto \mu(n, m) \in M$ are A-linear. The mapping μ is sometimes called the *multiplication*, or *multiplicative operation*, of the A-algebra. To see an example take $A = \mathbb{R}$, $M = \mathbb{R}^3$ and let $\mu : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ be the usual cross (or "vector") product $(v, w) \mapsto v \times w$.

When A is clear from context an A-algebra is simply called an *algebra*.

An A-algebra M with multiplication μ is: *commutative* if $\mu(m, n) = \mu(n, m)$ for all $m, n \in M$; *associative* if $\mu(\mu(m, n), p) = \mu(m, \mu(n, p))$ for all $m, n, p \in M$. The example ending the previous paragraph has neither property. A *unity*, or *multiplicative identity* for an A-algebra is an element $e \in M$ such that $\mu(e, m) = \mu(m, e) = m$ for all $m \in M$. An A-algebra which admits precisely one unity element is said to be an *A-algebra with unity*. $\mathbb{R}^3$ is an example of of an $\mathbb{R}$-algebra with no unity. To see an example of a commutative and associative A-algebra with unity let M be the polynomial algebra $A[x_1, \ldots, x_n]$ in n variables with "multiplication" meaning the usual multiplication of polynomials.

Our definition of an A-algebra is sufficiently general to include Lie algebras (which will be defined when needed), and this is why we have opted for this particular formulation. A more restrictive definition is often encountered, and since that definition is sufficient for our immediate purposes it is worth brief mention. Specifically, one frequently sees an "A-algebra" defined as a not necessarily commutative ring B together with a ring homomorphism $f : A \to B$ satisfying

$$\text{(i)} \qquad f(a)b = bf(a) \qquad \text{for all} \qquad a \in A \qquad \text{and all} \qquad b \in B.$$

To relate this to our definition simply note that we can give B the structure of an A-module by defining $a \cdot b := f(a)b$, $a \in A$, $b \in B$, whereupon condition (i) ensures that multiplication in B is A-bilinear. In other words, by means of the given ring homomorphism B can be viewed an A-algebra in our original sense, with multiplication being that already assumed in B. There are four important examples of A-algebras of this restricted type to keep in mind: (a) $1 \in A$ is the unity, and $f : \mathbb{Z} \to A$ is the ring homomorphism defined by

$$\text{(ii)} \qquad n \mapsto n \cdot 1 := \begin{cases} \sum_{j=1}^{n} 1 & \text{if} \quad n > 0 \\ 0 & \text{if} \quad n = 0 \\ \sum_{j=1}^{|n|} (-1) & \text{if} \quad n < 0\,; \end{cases}$$

(b) B is any (not necessarily commutative) extension ring of A with $f : A \to B$ given by inclusion, e.g., B is the polynomial algebra $A[x_1, \ldots, x_n]$ in n indeterminates for some $1 \leq n \in \mathbb{Z}$; (c) $\mathfrak{i} \subset A$ is an ideal and $f : A \to A/\mathfrak{i}$ is the canonical homomorphism; (d) B is the ring of $(n \times n)$-matrices

with entries in a commutative ring A, where again $1 \leq n \in \mathbb{Z}$, and f is the ring homomorphism

$$\text{(iii)} \qquad a \in A \mapsto \begin{pmatrix} a & 0 & 0 & \cdots & 0 \\ 0 & a & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & & & a & 0 \\ 0 & 0 & \cdots & 0 & a \end{pmatrix} \in B.$$

From Example (a) of the previous paragraph we see that any ring is a $\mathbb{Z}$-algebra. The range if the indicated homomorphism is the *image of* $\mathbb{Z}$ *in* A.

When A and the ring homomorphism $f : A \to B$ of the previous paragraph are understood it is customary to refer to B, rather than to f, as the A-algebra. We will follow this custom. The practice is no doubt familiar, at the very least, in the context of polynomial algebras. When a bit more precision is needed to avoid confusion we will indicate this situation by asserting that B is an A-algebra by virtue of the ring homomorphism $f : A \to B$.

[5]Suppose M and N are A-algebras (as initially defined in Footnote 4) with multiplicative operations μ_M and μ_N respectively. An A-module homomorphism $f : M \to N$ is an A-*algebra homomorphism* if

$$\text{(iv)} \qquad (f \circ \mu_M)(m_1, m_2) = f(\mu_M(m_1, m_2)) = \mu_N(f(m_1), f(m_2)) \qquad \text{for all} \qquad m_1, m_2 \in M.$$

For algebras of the restricted type the following definition suffices: an A-*algebra homomorphism* between A-algebras $f : A \to B$ and $g : A \to C$ is a ring homomorphism $h : B \to C$ such that the diagram

$$\text{(v)} \qquad \begin{array}{ccc} & & B \\ & \overset{f}{\nearrow} & \\ A & & \downarrow h \\ & \underset{g}{\searrow} & \\ & & C \end{array}$$

commutes. To verify that h is an A-linear module homomorphism, i.e., that this is a special case of the previous definition of an A-algebra homomorphism, simply note that for any $a \in A$ and $b \in B$ one has

$$\text{(vi)} \qquad \begin{aligned} h(a \cdot b) &= h(f(a)b) \\ &= h(f(a))h(b) \\ &= g(a)h(b) \\ &= a \cdot h(b). \end{aligned}$$

When A is understood an A-algebra homomorphism will be called an *algebra homomorphism*.

bra homomorphism, algebra isomorphism, integral domain, and field[6]. However, we make no assumptions about the existence of any rings, integral domains or algebras other than the usual ring $\mathbb{Z}$ of integers, the polynomial $\mathbb{Z}$-algebra $\mathbb{Z}[x]$ in a single indeterminate x, and matrix algebras with coefficients in $\mathbb{Z}$. Nor do we assume the existence of any fields, or of any modules other than the $\mathbb{Z}$-modules $\mathbb{Z}[x]$ and (the Cartesian products) $\mathbb{Z}$, $\mathbb{Z}^2$, $\mathbb{Z}^3$, ... On the other hand, with regard to the ring $\mathbb{Z}$ we assume knowledge of the Fundamental Theorem of Arithmetic (unique factorization of integers into primes), greatest common divisors and least common multiples of non-empty finite sets of integers, and relatively prime pairs of integers.

(a) A subset $\mathfrak{i} \subset A$ is an *ideal* if it is an A-module when addition and scalar multiplication are assumed induced from A, i.e., if $\mathfrak{i}$ is a (necessarily abelian) group under addition and if $ai \in \mathfrak{i}$ whenever $a \in A$ and $i \in \mathfrak{i}$. Examples: $\mathfrak{i} = A$ (all other ideals are *proper*); $\mathfrak{i} = \{0\}$ (the *zero ideal*); $\mathfrak{i}$ consists of all multiples of a fixed element $a \in A$, in which case $\mathfrak{i}$ is generally written as (a) or as aA. Ideals of this last form are called *principal*; specifically, (a) is the *principal ideal generated by* a. Note that $A = (1)$ and $\{0\} = (0)$: these ideals are *trivial*; all others (if any) are *non-trivial*.

Any ideal $\mathfrak{i} \subset A$ defines a relation $\sim$ on A by $a \sim b \Leftrightarrow a - b \in \mathfrak{i}$. This is an equivalence relation, as is easily checked. The equivalence class (i.e., coset) $[a] := a + \mathfrak{i}$ of an element $a \in A$ is called the *residue class of* a $(mod(ulo)\ \mathfrak{i})$, and the collection of these classes is denoted $A/\mathfrak{i}$ (read: A mod(ulo) I). Addition and multiplication on $A/\mathfrak{i}$ are well-defined by $[a] + [b] := [a + b]$ and $[a] \cdot [b] := [ab]$ respectively, $a, b \in A$, and the collection is thereby given the structure of a (commutative) ring with unity [1]; it is the *residue class ring, factor ring,* or *quotient ring, of* A *by* $\mathfrak{i}$. The mapping $\varphi : a \in A \to [a] \in A/\mathfrak{i}$ is easily seen to be a ring homomorphism, called the *canonical homomorphism*. Note that φ has kernel $\mathfrak{i}$. More generally, the kernel $\ker(f)$ of any ring homomorphism $f : A \to B$ is an ideal.

Examples :

(i) For any positive integer n the factor ring $\mathbb{Z}/n\mathbb{Z}$ is the *ring of integers modulo* n. It is a field if and only if n is a prime number, and as a result we suddenly have infinitely many fields at our disposal. Applying the canonical homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is called *reduction mod(ulo)*

[6]We take $1 \neq 0$ to be one of the defining conditions for both integral domains and fields. Moreover, we only consider commutative integral domains and commutative fields.

n; applying φ to an integer $a \in \mathbb{Z}$ is referred to as *reducing* a *mod*(*ulo*) n.

In the number-theoretical context equality of cosets $[a], [b] \in \mathbb{Z}/n\mathbb{Z}$ is commonly expressed as $a \equiv b \pmod{n}$ (read: a is congruent to b mod(ulo) n) rather than as $[a] = [b]$. We prefer the latter notation since it emphasizes the generality of the concept.

(ii) For any indeterminate x over the ring A the mapping $a \in A \mapsto [a] \in A[x]/(x) = A[x]/xA[x]$ is a ring isomorphism of A and the factor ring $A[x]/(x)$.

(iii) (Manufacturing roots of polynomials) Let $A[x]$ be as in (ii) and let $p \in A[x]$ be a non-constant polynomial with no roots in A. Then the composition $A \hookrightarrow A[x] \xrightarrow{\varphi} A[x]/(p)$ of the inclusion mapping $A \hookrightarrow A[x]$ with the canonical homomorphism $\varphi : A[x] \to A[x]/(p)$ is a ring embedding, allowing us to view A as a subring of $A[x]/(p)$, and $[x] \in A[x]/(p)$ is (by construction) a root of p. The practice is to identify A with the subring $\varphi(A) \subset A[x]/(p)$, for this reason one writes $\varphi(a)$ as a whenever $a \in A$.

To see a specific example take $A = \mathbb{Z}$ and $p = x^2 - 10$. We claim that a typical element of $\mathbb{Z}[x]/(p)$ can be written in the form $a + b\sqrt{10}$, wherein $a, b \in \mathbb{Z}$ and $\sqrt{10} := [x]$, and because of this fact one writes $\mathbb{Z}[x]/(x^2 - 10)$ as $\mathbb{Z}[\sqrt{10}\,]$. Indeed, first observe that $\sqrt{10} := [x]$ implies $[x]^2 = (\sqrt{10}\,)^2 = 10$, and as consequences we see that every positive even power of $[x]$ is in $\mathbb{Z}$ (in fact must be a power of 10) and that every positive odd power is an integer multiple of $\sqrt{10}$. Now simply note from the form $\sum_{j=0}^{n} a_j x^j$ of a typical polynomial in $\mathbb{Z}[x]$ that any element of $\mathbb{Z}[x]/(p)$ can be written as $\sum_{j=0}^{n} \varphi(a_j)(\varphi(x))^j = \sum_{j=1}^{n} a_j [x]^j$, which by the prior observation can be reduced to the form $a + b\sqrt{10}$ with $a, b \in \mathbb{Z}$. In this case the canonical homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}[\sqrt{10}]$ is given by $a \mapsto a + 0 \cdot \sqrt{10}$, which one would ordinarily express as $a \in \mathbb{Z} \mapsto a \in \mathbb{Z}[\sqrt{10}]$.

Rings with similar constructions are indicated with analogous notation, e.g., $\mathbb{Z}[i] = \mathbb{Z}[\sqrt{-1}\,]$ would refer to the ring $\mathbb{Z}[x]/(x^2 + 1)$, wherein each element can be written in the form $a + bi$ with $a, b \in \mathbb{Z}$ and $i := \sqrt{-1} := [x]$ (which guarantees that $i^2 = -1$). $\mathbb{Z}[i]$ is the *ring of Gaussian integers*. Here the canonical homomorphism is $a \in \mathbb{Z} \mapsto a + 0 \cdot i$ ($:= a \in \mathbb{Z}[i]$).

A fundamental example for algebraic number theory is the ring $\mathbb{Z}[\zeta] := \mathbb{Z}[x]/(x^2 + x + 1)$. One refers to $\zeta := [x]$ as a *primitive cube root of unity*:

from $x^3-1=(x-1)(x^2+x+1)$ one sees that $\zeta^3=1$. The typical element of $\mathbb{Z}[\zeta]$ can be written in the form $a+b\zeta,\ a,b\in\mathbb{Z}$, and the canonical homomorphism is $a\mapsto a+0\cdot\zeta$ $(:=a\in\mathbb{Z}[\zeta])$.

It is important to keep in mind that we are not regarding $\sqrt{10}$ as a real number, nor are we regarding i and ζ as complex numbers: the real and complex fields have yet to be constructed.

Application I: *The equation* $x^3+y^3=z^3$ *has no solutions* $x,y,z\in\mathbb{Z}$ *with* $xyz\neq 0$ *and none of* x, y *and* z *divisible by* 3. For suppose $a,b,c\in\mathbb{Z}$ satisfy

$$a^3+b^3=c^3$$

and $abc\neq 0$. Applying the canonical homomorphism $\varphi:\mathbb{Z}\to\mathbb{Z}/9\mathbb{Z}$ then gives

$$[a]^3+[b]^3=[c]^3$$

with $[a],[b],[c]\in\{[1],[2],[4],[5],[7],[8]\}$. But this cannot be: examine the possibilities using $[n]^3\in\{[1],-[1]\}$ for $n=1,2,4,5,7,8$.

Application : Problems II and III at the beginning of the section can be viewed as problems about multiplication in the ring $\mathbb{Z}[i]$ of Gaussian integers. Indeed, in that ring we can write x^2+y^2 as $(x+iy)(x-iy)$, and as a result we see that II is equivalent to: find all solutions of $(x+iy)(x-iy)=z^2$ with $x,y,z\in\mathbb{Z}$ and $xyz\neq 0$. Similarly, III is equivalent to: for which positive integers n can one find non-negative integers x,y such that $(x+iy)(x-iy)=n$?

(b) A subset $S\subset A$ containing 1 and closed under multiplication is *multiplicative*. Examples: $S:=A\backslash\{0\}$ when A is an integral domain; $S=\{a^n\}_{n=0}^{\infty}$ for an arbitrary but fixed element $a\in A$; $S:=\{0,1\}$.

Any multiplicative subset $S\subset A$ defines a relation $\sim$ on $A\times S$ by $(a_1,s_1)\sim(a_2,s_2)$ if and only if there is an element $s\in S$ such that $sa_1s_2=sa_2s_1$. This is again an equivalence relation; the verification is straightforward. Write the equivalence class of an element $(a,s)\in A\times S$ as a/s and denote the set of equivalence classes by $S^{-1}A$. Addition and multiplication are well-defined on $S^{-1}A$ by

$$a_1/s_1+a_2/s_2:=(a_1s_2+a_2s_1)/s_1s_2$$

and

$$(a_1/s_1)\cdot(a_2/s_2):=a_1a_2/s_1s_2$$

respectively, and the collection is thereby given the structure of a (commutative) ring with unity $1/1$; it is the *ring of fractions of* A *by* S. The mapping $\varphi : a \in A \mapsto a/1 \in S^{-1}A$ is easily seen to be a ring homomorphism, called the *canonical homomorphism*. Note that for any $s \in S$ the element $\varphi(s) = s/1 \in S^{-1}A$ is invertible (with inverse $1/s$).

When A *is an integral domain and* $S \subset A\backslash\{0\}$ *one has* $(a_1, s_1) \sim (a_2, s_2)$ *if and only if* $a_1 s_2 = a_2 s_1$. Argue as follows: by definition one has $(a_1, s_1) \sim (a_2, s_2)$ if and only if there is an element $s \in S$ such that $sa_1 s_2 = sa_2 s_1$, which in turn holds if and only if $s(a_1 s_2 - a_2 s_1) = 0$. However, since A is an integral domain and $s \neq 0$ this is equivalent to $a_1 s_2 = a_2 s_1$.

When A *is an integral domain and* $S \subset A\backslash\{0\}$ *the canonical homomorphism* $\varphi : A \to S^{-1}A$ *is an embedding.* Indeed, from the previous paragraph one has $\varphi(a_1) = \varphi(a_2) \Leftrightarrow a_1/1 = a_2/1 \Leftrightarrow a_1 \cdot 1 = a_2 \cdot 1 \Leftrightarrow a_1 = a_2$. In this context φ is used to identify A with the subdomain $\varphi(A) \subset S^{-1}A$. For example, one might replace the notation $\varphi : A \to S^{-1}A$ with $A \subset S^{-1}A$ and refer to the ring $S^{-1}A$ as an extension of A. Moreover, when $a \in A$ is identified with $\varphi(a) := a/1 \in S^{-1}A$ one would more likely write $a \in S^{-1}A$ than $a/1 \in S^{-1}A$.

In general one thinks of $\varphi : A \to S^{-1}A$ as "pushing" A into a ring stocked with inverses for the elements of S. Of course when φ fails to be an embedding one cannot take this intuitive viewpoint very seriously.

Examples:

(i) When $A = \mathbb{Z}$ and $S = \mathbb{Z}\backslash\{0\}$ the ring $S^{-1}\mathbb{Z}$ is a field; it is denoted $\mathbb{Q}$ and called the field of of *rational numbers*[7]. The canonical homomorphism is the standard embedding $n \in \mathbb{Z} \mapsto n/1 \in \mathbb{Q}$ used to regard $\mathbb{Z}$ as a subdomain of $\mathbb{Q}$.

(ii) More generally, when A is an integral domain and $S := A\backslash\{0\}$ the ring $S^{-1}A$ is a field, called "the" *quotient field* of A, and the canonical homomorphism is again a ring embedding. For example, when A is the polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ in n indeterminates the quotient field, denoted $\mathbb{Z}(x_1, \ldots, x_n)$, consists of all quotients of polynomials in $\mathbb{Z}[x_1, \ldots, x_n]$. The canonical homomorphism carries a polynomial $p \in \mathbb{Z}[x_1, \ldots, x_n]$ to the element $p/1 \in \mathbb{Z}(x_1, \ldots, x_n)$.

[7] As if we were unaware of this fact! Keep in mind that our presentation in this section is somewhat tongue-in-cheek.

(iii) Fix any non-zero element $b \in A$ and let $S := \{b^n\}_{n=0}^{\infty}$. Then $S^{-1}A$, which is generally written as A_b or $A[b^{-1}]$, consists of all quotients a/b^n with $0 \leq n \in \mathbb{Z}$ arbitrary. Alternatively, $A[b^{-1}]$ can be viewed as the collection of "polynomials $a_0 + a_1 b^{-1} + \cdots + a_n b^{-n}$ in b^{-1} with coefficients in A;" any such expression is easily reduced to the quotient form a/b^n.

(iv) When $S \subset A$ is multiplicative and contains 0 all pairs within $A \times S$ are equivalent and $S^{-1}A$ is the trivial ring 0 (i.e., the ring with only one element). In particular, in this context identities such as $0/1 = 1/0$ make sense.

By taking A to be a non-trivial ring we see that the canonical homomorphism $\varphi : A \to S^{-1}A$ need not be an embedding. In particular, our intuitive description of φ "pushing" A into another ring is a bit misleading, since we now see that collapsing can occur.

Application : *Suppose $a, b, c \in \mathbb{Z}$ and $a \neq 0$. Then the polynomial $ax^2+bx+c \in \mathbb{Z}[x]$ is reducible*[8] *in $\mathbb{Z}[x]$ if and only if b^2-4ac is the square of an integer.* First suppose the polynomial factors in $\mathbb{Z}[x]$ as $ax^2+bx+c = (dx+e)(fx+g) = dfx^2 + (dg+ef)x + eg$. Comparing coefficients gives $a = df$, $b = dg + ef$ and $c = eg$, hence

$$b^2 - 4ac = (dg+ef)^2 - 4dfeg = (dg-ef)^2,$$

and the forward assertion is thereby established. Now[9] assume $b^2 - 4ac = n^2$ for some integer n. Replacing $\mathbb{Z}[x]$ with the ring $\mathbb{Q}[x]$ we can then write

$$\begin{aligned}
ax^2 + bx + c &= a\left(x^2 + \tfrac{b}{a}x + \tfrac{b^2}{4a^2}\right) + c - \tfrac{b^2}{4a} \\
&= a\left(x + \tfrac{b}{2a}\right)^2 - \tfrac{b^2-4ac}{4a} \\
&= a\left(\left(x + \tfrac{b}{2a}\right)^2 - \tfrac{b^2-4ac}{4a^2}\right) \\
&= a\left(\left(x + \tfrac{b}{2a}\right)^2 - \tfrac{n^2}{4a^2}\right) \\
&= a\left(\left(x + \tfrac{b}{2a}\right) + \tfrac{n}{2a}\right)\left(\left(x + \tfrac{b}{2a}\right) - \tfrac{n}{2a}\right) \\
&= a\left(x + \tfrac{b+n}{2a}\right)\left(x + \tfrac{b-n}{2a}\right).
\end{aligned}$$

[8] By "reducible in $\mathbb{Z}[x]$" we mean: factors in the form $(dx+e)(fx+g)$, where $d, e, f, g \in \mathbb{Z}$.

[9] The converse is an easy consequence of the fact that a polynomial in $\mathbb{Z}[x]$ factors into polynomials in $\mathbb{Z}[x]$ of lower degrees if and only if it has a corresponding degree-preserving factorization in $\mathbb{Q}[x]$. That result, however, requires the Gauss Lemma. Our restricted context allows for a more elementary argument.

From $b^2 - n^2 = 4ac$ we see that b and n must have the same parity[10], and it follows easily that $b+n$ and $b-n$ must be even. In particular, the rational numbers $\frac{1}{2}(b \pm n)$ must be integers, and we can therefore write $ac = \frac{b+n}{2} \cdot \frac{b-n}{2}$ as an identity within $\mathbb{Z}$. By unique factorization any prime factor of a (assuming $a \neq \pm 1$) must be a prime factor of at least one of $\frac{b+n}{2}$ or $\frac{b-n}{2}$, and we can therefore write $a = a_1 a_2$, where $a_1 \in \mathbb{Z}$ is a factor of $\frac{b+n}{2}$ and $a_2 \in \mathbb{Z}$ is a factor of $\frac{b-n}{2}$ (with $a_1 = 1$ and/or $a_2 = 1$ being distinct possibilities). This gives $a_2 \cdot \frac{b+n}{2a} = \frac{b+n}{2a_1} \in \mathbb{Z}$ and $a_1 \cdot \frac{b-n}{2a} = \frac{b-n}{2a_2} \in \mathbb{Z}$, and the factorization $ax^2 + bx + c = (a_2 x + \frac{b+n}{2a_1})(a_1 x + \frac{b-n}{2a_2})$ is thereby seen to be in $\mathbb{Z}[x]$.

Constructions (a) and (b) are often combined. For example, one can first construct the field of rational numbers $\mathbb{Q}$ from $\mathbb{Z}$ as in (bi), and then take $A = \mathbb{Q}$ and $p = x^2 + 1$ in (aiii) to construct the ring $\mathbb{Q}[i]$.

We detailed the construction of the field of rational numbers for two reasons: to illustrate Kronecker's statement; and to fit that construction into a more general context. We have no intention of rigorously establishing the standard arithmetical and order properties of $\mathbb{Q}$; these are henceforth assumed familiar to readers. The same assumption applies to the elementary arithmetical properties of the auxiliary rings we have constructed. On the other hand, properties of these auxiliary rings which we regard as somewhat less than elementary will be established rigorously. This adjustment in attitude is illustrated in the proof of the next result, wherein we assume knowledge of the elementary ring properties of both $\mathbb{Q}$ and $\mathbb{Z}[i]$, e.g., the fact that the square of any rational number is a non-negative rational number.

Proposition 1.1 : *Let A denote either $\mathbb{Z}$ or $\mathbb{Q}$. Then any element of $A[i]$ can be written uniquely in the form $a + ib$, with $a, b \in A$. Moreover, $A[i]$ is an integral domain if $A = \mathbb{Z}$; it is a field if $A = \mathbb{Q}$.*

Proof : Since $\mathbb{Z}[i] \subset \mathbb{Q}[i]$ it suffices to prove the initial assertion when $A = \mathbb{Q}$. However, if $a + ib = c + id$ and $b \neq d$ then $i = (c-a)/(b-d) \in \mathbb{Q}$, which is impossible since $i^2 = -1 < 0$. Now simply note that $b = d \Rightarrow a = c$.

To prove the integral domain/field assertions it suffices, in view of the inclusion $\mathbb{Z}[i] \subset \mathbb{Q}[i]$, to prove that the latter is a field. But this is immediate from the observation that the inverse of a non-zero element $a + ib \in \mathbb{Q}[i]$ is given by $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2} \cdot i$. **q.e.d.**

[10] I.e., both must be odd or both must be even.

Application II : We use the preceding ideas to solve the Pythagorean triple problem discussed at the beginning of the section.

Theorem : *A triple of non-negative integers* (a, b, c) *satisfies the Diophantine equation* $x^2 + y^2 = z^2$ *if and only if there are integers* m, n *such that this triple is proportional to* $(m^2 - n^2,\ 2mn,\ m^2 + n^2)$.

Here "is proportional" means: has the form

(i) $\quad (a, b, c) = k(m^2 - n^2, 2mn, m^2 + n^2) := (k(m^2 - n^2), k \cdot 2 \cdot mn, k(m^2 + n^2))$

for some non-negative rational number k.

Since a triple (a, b, c) of integers satisfies $x^2 + y^2 = z^2$ if and only if this is the case when any or all of these integers is replaced by its negative, the restriction $c \geq 0$ in the statement of this result is inconsequential.

The formulation and proof of this theorem are adapted from [El], where the argument is related to Hilbert's Theorem 90.

Proof :

$\Rightarrow$ Suppose $(a, b, c) \in \mathbb{Z}^3$ satisfies the given equation.

If $c = 0$ the same must hold for a and b, and in that instance we can achieve the form given in (i) by choosing $k = 0$ and m and n arbitrarily. We therefore assume $c \neq 0$, in which case at least one of a and b must be non-zero.

If $b = 0$ then $a = c$ (since both are non-negative) and we can achieve (i) by taking $m = a$, $n = 0$ and $k = 1/a$.

If $a = 0$ then $b = c$ and we can achieve (i) by taking $m = n = c$ and and $k = 1/2c$.

We therefore assume $abc \neq 0$. Define integers m and n by

$$m := 2bc \qquad \text{and} \qquad n := 2c(c - a).$$

Then $0 \neq m + in \in \mathbb{Z}[i]$, and in $\mathbb{Q}(i)$ we have

$$\frac{a + ib}{c} = \frac{m + in}{m - in} = \frac{(m + in)(m + in)}{(m - in)(m + in)} = \frac{(m^2 - n^2) + 2imn}{m^2 + n^2}.$$

The result follows.

$\Leftarrow$ One sees from the identity $(m^2 - n^2)^2 + (2mn)^2 = (m^2 + n^2)^2$ that integers $a := m^2 - n^2$, $b := 2mn$ and $c := m^2 + n^2$ satisfy the equation $x^2 + y^2 = z^2$ for any choice of the integers m, n. Moreover, the same is obviously true of any multiple (ka, kb, kc) with $k \in \mathbb{Q}$ and $ka, kb, kc \in \mathbb{Z}$.

q.e.d.

2. Preliminaries on Divisibility

In this section we recall a few basic definitions and results from commutative ring theory. This is done primarily to establish notation. The results are used in later sections, generally without comment.

Throughout this section A *denotes a commutative ring with unity* 1*, and* $1 = 0$ *is allowed. The product of two elements* $a, b \in A$ *is written as* ab *or, when confusion might otherwise result, as* $a \cdot b$*.*

Let $a, b \in A$. We say that a *divides* b, and write $a|b$, if $b = ac$ for some element $c \in A$. We also read $a|b$ as: a is a *divisor* of b; b is a *multiple* of a; a is a *factor* of b; and b is *divisible* by a. The negation of $a|b$ is written $a \nmid b$, which one reads as a does not divide b. Elements $a, b \in A$ satisfying $a|b$ and $b|a$ are called *associates*. Examples: In $\mathbb{Z}$ we have $3|75$, since $75 = 3 \cdot 25$; in any field we have $a|b$ for any two non-zero elements a, b; in the polynomial ring $(\mathbb{Z}/6\mathbb{Z})[x]$ (one indeterminate, i.e., one variable) we have $([2]x + [3])|x$ becuase $([2]x + [3])([3]x + [2]) = x$; any element of any ring is an associate of itself; in the ring $\mathbb{Z}$ any integer n has n and $-n$ as associates, and no other associates; in the ring $\mathbb{Z}[x]/(5x)$ the elements $[x]$ and $[2x]$ are associates since $[2x] = [2][x]$ and $[x] = [8][2x]$.

Note that for all $a \in A$ one has $1|a$ and $a|a$. For elements $a, b, c \in A$ the conditions $a|b$ and $a|c$ are easily seen to imply $a|(b \pm c)$.

An element $a \in A$ is:

- a *unit*, or is *invertible*, if $a|1$;
- *prime* if a is a nonzero non-unit and $a|bc$ for some $b, c \in A$ implies either $a|b$ or $a|c$ (or both);
- *irreducible* if a is a nonzero non-unit and $a = bc$ for some $b, c \in A$ implies that one of b and c must be a unit;
- a *zero divisor* if a is non-zero and $ab = 0$ for some non-zero element $b \in A$.

When a is a unit there is, by definition, an element $b \in A$ such that $ab = 1$. This element is easily seen to be unique: it is the *inverse* of a and is denoted a^{-1}. Examples: The units of $\mathbb{Z}$ are ± 1; one has $1^{-1} = 1$ and $(-1)^{-1} = -1$. The units of $\mathbb{Z}/5\mathbb{Z}$ are $[1], [2], [3]$ and $[4]$; one has $[1]^{-1} = [1]$, $[2]^{-1} = [3]$, $[3]^{-1} = [2]$, and $[4]^{-1} = [4]$. In $\mathbb{Z}/6\mathbb{Z}$ the element $[2]$ is both prime and a zero divisor, but is not

irreducible. In $\mathbb{Z}[\sqrt{10}\,]$ the element 2 is irreducible but not prime[11]. Let K be a field and let $A := K[x]$ be the associated polynomial algebra in a single indeterminate: then $x \in A$ is irreducible.

Since $1|a$ for any $a \in A$ we see that a is a unit if and only if a is an associate of 1.

In the ring $\mathbb{Z}$ "prime number" is commonly understood to mean "positive prime number," although by the above definition the negative of any such prime would also qualify. The irreducibles in $\mathbb{Z}$ are the same as the primes (see Theorems 2.2(j) and 2.11(a)).

Proposition 2.1 : *Primes and irreducibles cannot divide units.*

Proof : Let $p \in A$ be prime or irreducible and let $u \in A$ be a unit. If $p|u$ then from $u|1$ we see that $p|1$, and p is therefore a unit. However, primes and irreducibles are, by definition, not units. **q.e.d.**

Recall that (a) and aA are used to denote the principal ideal generated by the element $a \in A$. The ring A is a *principal ideal domain*, or simply a PID, if A is an integral domain and all ideals are principal. Standard examples[12] of PIDs are the usual ring $\mathbb{Z}$ of integers and the polynomial rings $K[x]$ for any field K.

Theorem 2.2 :

(a) *Elements $a, b \in A$ are associates if and only if $(a) = (b)$.*

(b) *The collection of units of A is a group under multiplication.*

(c) *Suppose $a, b, c \in A$ and a and c are associates. Then $a|b$ if and only if $c|b$.*

(d) *Suppose $a, b, c \in A$ and a and c are associates. Then $b|a$ if and only if $b|c$.*

(e) *Any associate of a prime is prime.*

(f) *Any associate of an irreducible is irreducible.*

(g) *When $a, u \in A$ and u is a unit the elements a and ua are associates.*

(h) *When A is an integral domain elements $a, b \in A$ are associates if and only if $b = ua$ for some unit $u \in A$.*

[11] See, e.g., [Hun, Chapter III, §3, Exercise 1, p. 140].

[12] These particular examples of PIDs are assumed familiar to readers; the major steps in verifying this property are reviewed in Examples 2.9(a) and (b) and Theorem 2.11(a).

(i) *When* A *is an integral domain all primes are irreducible.*

(j) *When* A *is a* PID *the primes and irreducibles coincide.*

The group in (b) is the *group of units* of A and is denoted $A^\times$. Example: $\mathbb{Z}^\times = \{1, -1\}$.

Assertion (h) is false when the integral domain hypothesis is dropped. For example, in the ring $\mathbb{Z}[x]/(5x)$ the elements $[x]$ and $[2x]$ are associates, as was already noted in the paragraph where associates were defined, but neither is the product of the other with a unit (because the only units in this ring are $\pm[1]$ and $[2x] \neq [x]\cdot(\pm[1])$).

When K is a field and $A := K[x]$ is the associated polynomial algebra in a single indeterminate we have seen[13] that $A := K[x]$ is a PID, and we have also seen[14] that x is irreducible in this ring. It is then immediate from Theorem 2.2(j) that x is prime. More generally, any element of $K[x]$ of the form $k_1x + k_0$ with $k_0, k_1 \in K$ and $k_1 \neq 0$ is irreducible, and therefore prime: if there are polynomials $a, b \in K[x]$ satisfying $ab = k_1x+k_0$ then by the additivity of degrees under polynomial multiplication we can assume w.l.o.g. that $\deg a = 0$ and $\deg b = 1$, hence that $a \in K \setminus \{0\}$ (because $k_1 \neq 0$), and a is therefore a unit (because K is a field).

The argument of the previous paragraph can fail when K is not assumed a field, the problem being that additivity of degrees need not hold. Indeed, at the beginning of this section we noted that $([2]x + [3])([3]x + [2]) = x$ in $(\mathbb{Z}/6\mathbb{Z})[x]$.

Proof : See, e.g., [Hun, Chapter III, §3, pp. 135-137]. **q.e.d.**

Units and irreducibles are most easily detected when A comes equipped with a *multiplicative norm*, i.e., a non-constant function $N : A \to \mathbb{Z}$ such that

$$N(ab) = N(a)N(b) \qquad \text{for all} \qquad a, b \in A. \tag{2.3}$$

We see directly from this definition that

$$a|b \qquad \Rightarrow \qquad N(a)|N(b), \tag{2.4}$$

[13] Immediately before the statement of Theorem 2.2.

[14] Three paragraphs above the statement of Proposition 2.1.

Examples 2.5 :

(a) The usual absolute value function $n \in \mathbb{Z} \mapsto |n| \in \mathbb{N}$ is a multiplicative norm.

(b) A multiplicative norm on the ring $\mathbb{Z}[i] := \{ a + ib \ : \ a, b \in \mathbb{Z} \}$ is defined by $a + ib \mapsto a^2 + b^2$. Proposition 1.1 is needed to ensure that this function is well-defined.

(c) The functions $a \in A \mapsto 1 \in \mathbb{Z}$ and $a \in A \mapsto 0 \in \mathbb{Z}$ satisfy (2.3), but are not multiplicative norms (since they are constant functions).

For the norms given in Examples 2.5(a) and (b) one checks easily that

$$(2.6) \qquad \begin{cases} \text{(a)} & N(1) = 1, \ \text{ and that} \\ \text{(b)} & N(a) = \pm 1 \ \text{ if and only if } \ a \ \text{ is a unit.} \end{cases}$$

Note that (a) and (2.3) imply $N(a) = \pm 1$ for all units $a \in A$; what is important about (b) is the converse. Easy application: the group of units $\mathbb{Z}[i]^\times$ of the Gaussian integers is $\{1, i, -1, -i\}$.

Proposition 2.7 : *Suppose $N : A \to \mathbb{Z}$ is a multiplicative norm satisfying the two conditions in* (2.6). *Then:*

(a) *any element $a \in A$ such that $N(a) \in \mathbb{Z}$ is prime must be irreducible; and*

(b) *when A is a* PID *any such element must be prime.*

Application: the Gaussian integers $1 - i$ and $1 + i$, are prime. In fact they are associate primes: one has $1 + i = i(1 - i)$, and $i \in \mathbb{Z}[i]$ is a unit (recall Theorem 2.2(h)).

Proof :

(a) Suppose $a, b, c \in A$ satisfy $a = bc$ and $N(a) \in \mathbb{Z}$ is prime. Then from (2.3) we see that one of $N(b)$ and $N(c)$ must have value 1 or -1, hence one of b and c must be a unit.

(b) By (a) and Theorem 2.2(j).

q.e.d.

We can get better results about units when the ring A admits a *Euclidean norm*, i.e., a function $\epsilon : A\backslash\{0\} \to \mathbb{Z}$ such that for all $a \in A\backslash\{0\}$ one has:

(a) $\epsilon(a) \geq 0$;

(b) $\epsilon(ab) \geq \epsilon(a)$ whenever $b \in A$ and $ab \neq 0$; and

(c) for each non-divisor $b \in A\backslash\{0\}$ of a there are (not necessarily unique) elements $c, r \in A\backslash\{0\}$ such that

$$a = cb + r \qquad \text{and} \qquad \epsilon(r) < \epsilon(b) \qquad \text{if} \qquad r \neq 0. \tag{2.8}$$

One refers to r as a "remainder" after "dividing a by b." Euclidean norms are also called *norms* and occasionally *height functions*[15].

A ring (*resp.* integral domain) with a Euclidean norm is called a *Euclidean ring* (resp. *Euclidean domain*[16]). Arithmetic in Euclidean domains closely parallels ordinary arithmetic in the usual ring of integers $\mathbb{Z}$.

Examples 2.9 :

(a) The usual absolute value function $n \in \mathbb{Z}\backslash\{0\} \mapsto |n|$ is a Euclidean norm on the ring of integers $\mathbb{Z}$. In particular, $\mathbb{Z}$ is a Euclidean domain. This Euclidean norm is always assumed unless specifically stated to the contrary.

(b) When K is a field and x is a single indeterminate over K the function $p \in K[x]\backslash\{0\} \mapsto \deg(p) \in \mathbb{Z}$ is a Euclidean norm on the polynomial ring $K[x]$. In particular, $K[x]$ is a Euclidean domain[17]. This *degree Euclidean norm* is always assumed unless specifically stated to the contrary[18].

(c) The function $p \in \mathbb{Z}[x]\backslash\{0\} \mapsto \deg(p)$ is <u>not</u> a Euclidean norm on the polynomial ring $\mathbb{Z}[x]$. For example, when $a = 3x^2 + 1$ and $b = 2x + 1$ we have $b \nmid a$, but there are no polynomials $c, d \in \mathbb{Z}[x]$ such that (2.8) holds.

(d) Any field K is a Euclidean domain; the function $k \in K\backslash\{0\} \to 0 \in \mathbb{Z}$ provides a Euclidean norm.

[15] See, e.g., [Koc, pp. 22], where the definition is a non-equivalent variation of what we have given.

[16] Readers are warned that for some authors "Euclidean ring" means what we have defined as a Euclidean domain.

[17] This is assumed familiar to readers.

[18] The degree of the zero polynomial of any polynomial ring is not defined.

(e) A Euclidean norm ϵ on $\mathbb{Z}/4\mathbb{Z}$ is defined by $[1] \mapsto 1$, $[2] \mapsto 2$ and $[3] \mapsto 1$. $\mathbb{Z}/4\mathbb{Z}$ thereby becomes an example of a Euclidean ring which is not a Euclidean domain[19]. One can illustrate non-uniqueness in (2.8) with this example, e.g., $[1] = [0][2] + [1] = [3][2] + [3]$.

(f) The multiplicative norm $a + ib \mapsto a^2 + b^2$ is a Euclidean norm on the ring of Gaussian integers[20].

We record a few elementary implications of the definition of a Euclidean norm.

Proposition 2.10 : *When A is a Euclidean ring with Euclidean norm ϵ the following statements hold.*

(a) *For any $0 \neq a \in A$ one has $\epsilon(1) \leq \epsilon(a)$.*

(b) *An element $a \in A$ is a unit if and only if $\epsilon(a) = \epsilon(1)$.*

(c) *A non-zero divisor $a \in A$ is a unit if and only if $\epsilon(ab) = \epsilon(b)$ for all $0 \neq b \in A$.*

(d) *A non-zero divisor $b \in A$ is a non-unit if and only if $\epsilon(a) < \epsilon(ab)$ for all non-zero divisors $a \in A$.*

(e) *If $a, b \in A$ are non-zero then $a|b \Rightarrow \epsilon(a) \leq \epsilon(b)$.*

(f) *When $a, b \in A$ are non-zero associates one has $\epsilon(a) = \epsilon(b)$.*

(g) *For any $0 \neq b \in A$ one has $\epsilon(b) = \epsilon(-b)$.*

The converse of (f) is false, e.g., $x^2 + x + 1$ and x^2 have the same degree, hence the same Euclidean norm in $\mathbb{Q}[x]$ (assuming the norm of Example 2.9(b)), but one sees easily that the polynomials are not associates.

Proof :

(a) $\epsilon(a) = \epsilon(1 \cdot a) \geq \epsilon(1)$.

(b) $1 = aa^{-1} \Rightarrow \epsilon(1) = \epsilon(aa^{-1}) \geq \epsilon(a) \geq \epsilon(1)$ (the last inequality by (a)), and therefore $\epsilon(a) = \epsilon(1)$ when a is a unit.

[19] **Proof :** Condition (a) in the definition of a Euclidean norm is evident, and, since the ring has only four elements, Condition (b) can be verified simply by checking all possible products. As for (c), the units of $\mathbb{Z}/4\mathbb{Z}$ are $[1]$ and $[3]$, and these divide all elements; as a consequence the only non-zero element of this ring which is not a divisor of some other element is $[2]$, and one checks from the multiplication table that this divides neither $[1]$ nor $[3]$. Condition (c) is then seen from $[1] = [0][2] + [1]$ and $[3] = [0][2] + [3]$. **q.e.d.**

[20] See, e.g., [Fral, Part IX, §47, Theorem 47.4, p. 408].

To prove the converse suppose $0 \neq a \in A$ satisfies $\epsilon(a) = \epsilon(1)$, but that a is not a unit. From Condition (c) in the definition of a Euclidean ring we can then choose $c, r \in A$ such that $1 = ca + r$ and $\epsilon(r) < \epsilon(a)$, hence $\epsilon(r) < \epsilon(1)$. This contradicts assertion (a).

(c) $\Rightarrow$: $\epsilon(b) = \epsilon(1 \cdot b) = \epsilon(a^{-1}a \cdot b) = \epsilon(a^{-1} \cdot ab) \geq \epsilon(ab) \geq \epsilon(b) \Rightarrow \epsilon(ab) = \epsilon(b)$.

$\Leftarrow$: Taking $b = 1$ we see that $\epsilon(a) = \epsilon(1)$, whereupon the result follows from assertion (b).

(d) $\Rightarrow$: If $\epsilon(a) = \epsilon(ab)$ for some non-zero divisor $a \in A$, write $a = ab \cdot c + r$, where either (i) $r = 0$ or (ii) $\epsilon(r) < \epsilon(ab) = \epsilon(a)$. If (i) holds then $a(1 - bc) = 0$, and since a is not a zero divisor this forces $bc = 1$, contradicting the assumption that b is not a unit. If (ii) holds then from $a(1-bc) = r$ we see that $\epsilon(a) \leq \epsilon(a(1-bc)) = \epsilon(r) < \epsilon(a)$, and we again have a contradiction.

$\Leftarrow$: Immediate from (c).

(e) By assumption there is a $c \in A$ such that $b = ac$, and therefore $\epsilon(b) = \epsilon(ac) \geq \epsilon(a)$.

(f) Immediate from (e).

(g) The elements b and $-b = (-1)b$ are associates; the result is therefore immediate from (f).

q.e.d.

The ring A is a *unique factorization domain*, or simply a UFD, if:

- A is an integral domain;
- every non-zero non-unit $a \in A$ has at least one "factorization into irreducibles", i.e., a factorization $a = \prod_{j=1}^{n} p_j$ into finitely many s irreducibles (with $n \geq 1$); and
- if $a = \prod_{k=1}^{m} q_k$ is a second such factorization of the same element one must have $m = n$ and there must be a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that p_i and $q_{\sigma(i)}$ are associates for $i = 1, 2, \ldots, n$.

Examples of factorizations into irreducibles are given by: $15 = (-3)(-5)$ in $\mathbb{Z}$; $2 = (1 + i)(1 - i)$ in $\mathbb{Z}[i]$; $2x^4 - 39x^3 + 231x^2 - 245x - 1029 = (2x + 3)(x - 7)^3$ in $\mathbb{Q}[x]$.

Fundamental examples of UFDs arise from (a) and (b) of the following result.

Theorem 2.11 :

(a) *Every Euclidean domain is a* PID*.*

(b) *Every* PID *is a* UFD*.*

(c) *In a* UFD *the primes and irreducibles coincide.*

We have seen that $\mathbb{Z}$, the polynomial ring $K[x]$ for any field K, and the ring of Gaussian integers are Euclidean domains, hence PIDs, and it follows that all provide examples of UFDs. An example of a UFD which is not a PID is given by[21] $\mathbb{Z}[x]$. An example of an integral domain which is not a UFD is given by $\mathbb{Z}[\sqrt{10}]$: in this integral domain one can factor any non-zero non-unit into a finite product of irreducibles, but the factorization is not unique in the sense defined above[22].

Proof : See, e.g., [Hun, Chapter III, §3, Theorems 3.9 and 3.7, pp. 138-9, and the Remark on p. 137]. **q.e.d.**

Corollary 2.12 : *Suppose A is a* UFD *and $N : A \to \mathbb{Z}$ is a multiplicative norm satisfying the two conditions in* (2.6)*. Then any element $a \in A$ such that $N(a) \in \mathbb{Z}$ is prime must also be prime.*

Proof : By (c) and Proposition 2.7(a). **q.e.d.**

The relation of being associates is an equivalence relation on A, and from Theorem 2.2(e) and (f) we see that this partitions the primes and irreducibles into equivalence classes. A collection of primes (*resp.* irreducibles) with precisely one element in each equivalence class is said to be a set of *representatives of the primes* (resp. *of the irreducibles*). Examples: (i) In the ring $\mathbb{Z}$ the set $P \subset \mathbb{Z}^+$ of positive primes constitute a set of representatives of the prime numbers. Indeed, as was noted immediately before Proposition 2.1, "prime numbers" in $\mathbb{Z}$ generally refer to "positive prime numbers." (ii) In the ring $\mathbb{Z}[i]$ of Gaussian integers we have seen[23] that the primes $1 + i$ and $1 - i$ are associates; hence both cannot be contained in a set of representatives of the primes.

Working with a set of representatives greatly simplifies statements and proofs of results about UFDs, e.g., if P is a set of representatives of the primes one can refer to unique factorization into elements of P without the usual qualification regarding associates. However, one must bring units into the picture.

[21] See, e.g., [Hun, Chapter III, §6, Theorem 6.14 and Exercise 1, pp. 164-5].

[22] See, e.g., [Hun, Chapter III, §3, Exercise 4, p. 140].

[23] Immediately following the statement of Proposition 2.7.

Proposition 2.13 : *Suppose* A *is a* UFD *and* $P \subset A$ *is a set of representatives of the primes. Then every non-zero non-unit* $a \in A$ *admits a factorization*

(i) $$a = u \cdot \prod_{i=1}^{m} p_i^{n_i}$$

in which u *is a unit,* $0 < m \in \mathbb{Z}$*, the* p_i *are distinct elements of* P*, and* $n_i \in \mathbb{Z}^+$ *for* $i = 1, 2, \ldots, n$*. The factorization is unique up to the order of the factors* $p_i^{n_i}$*.*

When P is understood it is customary to refer to (i) as "the prime factorization" of a, and we will follow this custom.

Examples: In the usual ring $\mathbb{Z}$ of integers $-1125 = (-1) \cdot 3^2 \cdot 5^3$ is the prime factorization of -1125 assuming the usual choice $\{2, 3, 5, 7, 11, \ldots\}$ for P. Similarly, $1125 = 3^2 \cdot 5^3$ is the prime factorization of 1125 (take $u = 1$). In the ring $\mathbb{Z}[i]$ of Gaussian integers $2 = i(1-i)^2$ is the prime factorization as in (i), assuming $1-i \in P$. (In particular, 2 is not a prime of this ring.) In the ring $\mathbb{Q}[x]$ (one indeterminate) $2x^4 - 39x^3 + 231x^2 - 245x - 1029 = 2(x + \frac{3}{2})(x - 7)^3$ is a prime factorization as in (i) (because 2 is a unit of $\mathbb{Q}$), assuming $x + \frac{3}{2}$, $x - 7 \in P$.

Proof : Since A is a UFD the irreducibles and primes coincide (Theorem 2.11(c)), and as a result there must be a non-empty collection of primes $q_1, q_2, \ldots, q_s \in A$ such that

(ii) $$a = \prod_{j=1}^{s} q_j.$$

Since P is a set of representatives of the primes each q_i must be an associate of a unique $p_i \in P$, and by Theorem 2.2(h) there must be a unit $u_j \in A$ such that $q_j = u_j p_j$, $j = 1, 2, \ldots, s$. Since the collection $A^\times$ of units forms a group the element $u := \prod_{j=1}^{s} u_j^{n_j}$ is again a unit. We obtain (i) by replacing each q_j in (ii) with $u_j p_j$ and then relabeling the p_j, if necessary, so that the n_j repetitions of any particular p_j are given the same subscript.

To prove uniqueness suppose $a = v \cdot \prod_{k=1}^{t} r_k^{s_k}$ is a second factorization having the asserted properties. Define $w := u^{-1}v \in A^\times$. Then wr_t is prime (by (g) and (e) of Theorem 2.2), and both sides of

(iii) $$\prod_{i=1}^{m} p_i^{n_i} = \prod_{k=1}^{t-s_t} r_k^{s_k} \cdot r_t^{s_t - 1} \cdot w r_t$$

can therefore be regarded as factorizations of the left-hand side into primes. It follows from unique factorization that $m = t$, and that each r_k for $1 \leq k < m$ must be an associate of some p_i. Since $r_k, p_i \in P$ and P is a set of representatives of the primes

this forces $r_k = p_i$ and $s_k = n_i$. Similarly, when $s_t > 1$ we must have $r_t = p_j$ for some $1 \leq j \leq n$. Since A is an integral domain (iii) can thereby be reduced, by repeated cancellation, to

$$p_i = wr_m \qquad \text{for some} \qquad 1 \leq i \leq m.$$

From Theorem 2.2(g) we conclude that p_i and r_t are associates, and since both are in P this forces $w = 1$ and $r_t = p_i$. Uniqueness follows. **q.e.d.**

Let $B \subset A$ be a non-empty subset. An element $c \in A$ is a *common divisor* of B if $b \in B \Rightarrow c|b$, and such a c is a *greatest common divisor* of B if c is divisible by all common divisors of B. The elements of B are *relatively prime* if 1 is a greatest common divisor of this set.

Proposition 2.14 : *Suppose A is a* UFD *and $\emptyset \neq B \subset A$. Then the following assertions hold.*

(a) *There is a non-unit common divisor of B if and only if this set admits a common divisor which is a prime.*

(b) *Suppose $P \subset A$ is a set of representatives of the primes. The elements of B are relatively prime if and only if there is no common divisor $p \in P$ of B.*

(c) *B admits a greatest common divisor.*

Proof :

(a) If $c \in A$ is a non-zero common divisor of B, and if $c = u \prod_{i=1}^{m} p^{n_i}$ a factorization of c as in Proposition 2.13, then any p_i is a common divisor of B which is prime.

The converse assertion is obvious.

(b) If 1 is a greatest common divisor of B then any common divisor of B divides 1, hence is a unit, hence cannot be prime.

Conversely, if there is no common divisor of B in P then by Proposition 2.13 any common divisor must be a unit, and therefore an associate of 1. The result then follows from Theorem 2.2(c).

(c) See, e.g., [Hun, Chapter III, §3, Theorem 3.11, p. 140].

q.e.d.

3. Ideal Arithmetic

The following observation enables one to convert divisibility questions into questions about containment of ideals (and vice-versa in the case of a PID).

Theorem 3.1 : *For any elements* $a, b \in A$ *the following statements are equivalent:*

(a) $a|b$*;*

(b) $b \in (a)$*; and*

(c) $(b) \subset (a)$*.*

Proof : Immediate from the definition of the principal ideal (c) generated by an element $c \in A$, i.e., $(c) := \{\, cd \,:\, d \in A \,\}$. **q.e.d.**

Let $\mathfrak{i}, \mathfrak{j} \subset A$ be ideals. Define their *sum* $\mathfrak{i} + \mathfrak{j}$ and *product* $\mathfrak{ij}$ by

$$\mathfrak{i} + \mathfrak{j} := \{\, a + b \in A \,:\, a \in \mathfrak{i},\ b \in \mathfrak{j} \,\} \tag{3.2}$$

and

$$\mathfrak{ij} := \{\, \textstyle\sum_{k=1}^{n} a_k b_k \,:\, a_k \in \mathfrak{i},\ b_k \in \mathfrak{j} \,\} \tag{3.3}$$

respectively, where in (3.3) the positive integer n is unrestricted. One checks easily that $\mathfrak{i} + \mathfrak{j}$ and $\mathfrak{ij}$ are ideals: $\mathfrak{i} + \mathfrak{j}$ is the smallest ideal containing $\mathfrak{i}$ and $\mathfrak{j}$, and $\mathfrak{ij}$ is the smallest ideal containing all products ab with $a \in \mathfrak{i}$ and $b \in \mathfrak{j}$. The question of immediate concern is: in familiar contexts how closely does this "ideal arithmetic" mimic ordinary addition and multiplication?

In the case of addition the answer is "not very well." For example, in the ring $\mathbb{Z}$ we might at first hope that $(2) + (3) = (5)$, but from $1 = -2 + 3 \in (2) + (3)$ we see that $(2) + (3) = (1) = \mathbb{Z}$.

However, in the case of multiplication the analogies are quite striking, particularly when A is a PID. Indeed, in the PID case we can write $\mathfrak{i} = (a)$ and $\mathfrak{j} = (b)$, and one sees easily that $\mathfrak{ij} = (ab)$. It follows by induction that

$$\textstyle\prod_{k=1}^{n} (a_k) = (\prod_{k=1}^{n} a_k), \tag{3.4}$$

and multiplication of ideals can therefore be achieved through the multiplication of generators. One refers to the ideals (a_k) in (3.4) as the *factors* of the ideal $(\prod a_k)$.

Are there analogues of primes and irreducibles? Indeed so.

A proper ideal $\mathfrak{i} \subset A$ is: *prime* if the residue class ring $A/\mathfrak{i}$ is an integral domain; *maximal* if $A/\mathfrak{i}$ is a field. Since fields are integral domains, every maximal ideal is prime. Equivalent definitions are: $\mathfrak{i}$ is prime if and if $ab \in \mathfrak{i}$ with $a, b \in A$ implies $a \in \mathfrak{i}$ or $b \in \mathfrak{i}$ (or both); $\mathfrak{i}$ is maximal if and only if for any ideal $\mathfrak{j} \subset A$ satisfying $\mathfrak{i} \subset \mathfrak{j}$ one has either $\mathfrak{i} = \mathfrak{j}$ or $\mathfrak{j} = A$. Examples: The ideal $(x) \subset \mathbb{Z}[x]$ is prime since $\mathbb{Z}[x]/(x) \simeq \mathbb{Z}$ is an integral domain; the ideal $(x) \subset \mathbb{Q}[x]$ is maximal since $\mathbb{Q}[x]/(x) \simeq \mathbb{Q}$ is a field; the ideal $(x^2+1) \subset \mathbb{Z}[x]$ is maximal since $\mathbb{Z}[x]/(x^2+1) \simeq \mathbb{Z}[i]$ is a field.

Theorem 3.5 :

(a) *A non-zero element $p \in A$ is prime if and only if (p) is a prime ideal.*

(b) *When A is an integral domain a non-zero element $p \in A$ is irreducible if and only if (p) is maximal* (w.r.t. inclusion) *among all proper principal ideals.*

(c) *When A is a* PID *the non-zero prime ideals and maximal ideals coincide.*

We have seen above that the principal ideal $(x) \subset \mathbb{Z}[x]$ (one indeterminate) is prime. Theorem 3.5(a) therefore offers a second proof[24] that $x \in \mathbb{Z}[x]$ is prime. We have also seen that the principal ideal $(x^2 + 1) \subset \mathbb{Z}[x]$ is maximal, and therefore prime. It now follows from Theorem 3.5(a) that $x^2 + 1 \in \mathbb{Z}[x]$ is prime.

Proof : See, e.g., [Hun, Chapter III, §3, pp. 135-137]. **q.e.d.**

Theorem 3.6 : *Let $P \subset \mathbb{Z}^+$ denote the collection of positive prime numbers. Then the mapping $p \in P \mapsto (p) \subset \mathbb{Z}$ is a bijection between P and the collection of non-zero prime ideals of $\mathbb{Z}$.*

Proof : The bijectivity of the mapping of Theorem 3.5(a) is an easy consequence of (a) and (e) of Proposition 2.2. **q.e.d.**

We have labeled the previous result as a theorem because of its fundamental importance to algebraic number theory. Indeed, the bijection enables one to generalize prime numbers by means of prime ideals. (Other venues for generalization are discrete valuations on $\mathbb{Z}$ and discrete valuation subrings of $\mathbb{Q}$.)

Proposition 3.7 : *Any non-zero proper ideal in a* PID *can be written uniquely, up to the order of the factors, as a product of prime ideals.*

[24] The first proof appears in the discussion following the statement of Theorem 2.2.

Note the absence of the annoying qualifications regarding associates which appear when formulating unique factorization in terms of ring elements. (The same qualifications mentioned in the paragraph preceding the statement of Proposition 2.13.) This fact alone suggests that the concept might better be formulated in terms of ideals, and in algebraic number theory this is precisely what one does. In that subject PIDs are subsumed by "Dedekind domains."

Proof : Any non-zero proper ideal has the form (a) for some non-zero non-unit a, and a admits a factorization $a = \prod_{j=1}^{n} p_j^{n_j}$ into pairwise non-associate irreducibles (by (b) of Theorem 2.11). By Theorem 2.2(j) the p_j are prime, by Theorem 3.5 the same holds for the ideals $(p_j) \subset A$, and from (3.4) we conclude that a factorization of (a) into prime ideals is given by $\prod_j (p_j)^{n_j}$.

To establish uniqueness rewrite the factorization $a = \prod_j p_j^{n_j}$ of the previous paragraph as $\prod_{j=1}^{s} p_j$, i.e., reduce all exponents to 1 by allowing p_i and p_j to be associates when $i \neq j$. Suppose $(a) = \prod_{k=1}^{t} (q_k)$ is any factorization of (a) into prime ideals. From (3.4) we have $(a) = (\prod_k q_k)$, we conclude from Theorem 2.2(a) that a and $\prod_k q_k$ are associates, whence from Theorem 2.2(h) that $a = u \prod_k q_k$ for some unit $u \in A$. If $u \neq 1$ replace q_1 by (the prime) uq_1 so as to write this last prime factorization as $a = \prod_{k=1}^{t} q_k$. By unique factorization we have $s = t$ and the existence of a permutation $\sigma : \{1, 2, \ldots, s\} \to \{1, 2, \ldots, s\}$ such that p_i and $q_{\sigma(i)}$ are associates for all $1 \leq i \leq n$. It follows from Theorem 2.2(a) that $(p_i) = (q_{\sigma(i)})$, and uniqueness is thereby established. **q.e.d.**

4. Derivations and Semi-Derivations

In this section A *denotes a not necessarily commutative ring with unity* 1*, and* $1 = 0$ *is again allowed.*

A mapping $\delta : a \in A \mapsto a' \in A$ is a *derivation* (on A) if for all $a, b \in A$ one has

$$(a+b)' = a' + b' \tag{4.1}$$

and

$$(a \cdot b)' = a \cdot b' + a' \cdot b. \tag{4.2}$$

One also writes a' as δa, or as $\delta(a)$ when confusion might otherwise result. Condition (4.1) is *additivity*; it is the assertion that δ is an additive group homomorphism. Condition (4.2) is the *Leibniz rule* or *product rule*. A mapping $\delta : a \in A \mapsto a' \in A$ satisfying (4.2) but not necessarily (4.1) is a *semi-derivation* (on A). (In particular, any derivation is a semi-derivation.)

When $\delta : A \to A$ is a semi-derivation or derivation and $a, b \in A$ satisfy $a' = b$ one often refers to b as the *derivative* of a and to a as a *primitive* of b.

Examples 4.3 :

(a) The primary example of a derivation is ordinary differentiation of polynomials: $A = R[x]$ for some commutative ring R and $\delta : R[x] \to R[x]$ is defined by $\sum_j r_j x^j \mapsto \sum_j j r_j x^{j-1}$. Additivity is clear, and the Leibniz rule is seen from

$$\begin{aligned}
\delta(\sum_i r_i x^i \sum_j s_j x^j) &= \delta(\sum_{ij} r_i s_j x^{i+j}) \\
= \sum_{ij} (i+j) r_i s_j x^{i+j-1} &= \sum_{ij} \left(r_i s_j j x^{(j-1)+i} + r_i i s_j x^{j+(i-1)} \right) \\
= (\sum_i r_i x^i)(\sum_j j s_j x^{j-1}) &+ (\sum_i i r_i x^{i-1})(\sum_j s_j x^j) \\
= (\sum_i r_i x^i)\delta(\sum_j s_j x^j) &+ \delta(\sum_i r_i x^i)(\sum_j s_j x^j).
\end{aligned}$$

δ is the *usual* or *standard derivation* on $R[x]$. When $\sum_j r_j x^j$ is expressed as $p(x)$ it is customary to write δ as $\frac{d}{dx}$ and $\delta(\sum_j r_j x^j)$ as $\frac{d}{dx}p(x)$ or $p'(x)$.

Examples: For $R = \mathbb{Q}$ and $a = x^4 - 3x^2 + 1 \in R[x]$ we have $(x^4 - 3x^2 + 1)' = 4x^3 - 6x$; for $R = \mathbb{Z}/5\mathbb{Z}$ and $a = x^6 - [7]x^5 + [3] \in R[x]$ we have $(x^6 - [7]x^5 + [3])' = 6x^5 - 5[7]x^4 + [0] = [1]x^5 - [35]x^4 = x^5 - [0]x^4 = x^5$.

(b) Suppose A is a UFD, $P \subset A$ is a set of representatives of the primes, and $p \in P$. Then by Proposition 2.13 any non-zero element $a \in A$ can be written uniquely in the form

$$a = bp^{\nu_p(a)}, \qquad \text{where} \qquad b \in A, \qquad p \not| \, b, \qquad \text{and} \qquad \nu_p(a) \in \mathbb{N}.$$

The integer $\nu_p(a)$ is the *p-adic value* or *p-adic valuation* of a. Define $\delta_p : A \to A$ by

$$\text{(i)} \qquad \delta_p : a \mapsto a' := \begin{cases} 0 & \text{if } a \neq 0 \text{ and } \nu_p(a) = 0, \\ \nu_p(a) b p^{\nu_p(a)-1} & \text{if } a \neq 0 \text{ and } \nu_p(a) \geq 1, \\ 0 & \text{if } a = 0. \end{cases}$$

Then δ_p is a semi-derivation on A which we call the *p-adic semi-derivation*[25]. (Verification of (4.2) is straightforward.)

To see specific examples take $A = \mathbb{Z}$, $P \subset \mathbb{Z}^+$ the collection of positive primes, $p = 5$ and $a = -75$. Then from the prime factorization $-75 = -3 \cdot 5^2$ we see that $b = -3$ and $\nu_5(-75) = 2$, hence that $\delta_5(-75) = (-75)' = 2 \cdot (-3) \cdot 5 = -30$. From the same prime factorization we see that $\delta_3(-75) = -25$ and $\delta_p(-75) = 0$ for all $p \in P \setminus \{3, 5\}$. From $1 = \delta_5(5) = \delta_5(2 + 3) \neq 0 + 0 = \delta_5(2) + \delta_5(3)$ we see that (4.1) can fail, i.e., that δ_p need not be a derivation.

For an example of different nature take $A = \mathbb{Z}[x]$ and let P include $p := x^2 + 1$. (We have previously noted[26] that p is prime.) Then from the prime factorization

$$x^8 - 8x^7 + 19x^6 - 24x^5 + 51x^4 - 24x^3 + 49x^2 - 8x + 16 = (x - 4)^2(x^2 + 1)^3$$

we see that

$$\begin{gathered} \delta_p(x^8 - 8x^7 + 19x^6 - 24x^5 + 51x^4 - 24x^3 + 49x^2 - 8x + 16) \\ = 3(x - 4)^2(x^2 + 1)^2 = 3x^6 + 54x^4 + 99x^2 - 24x^5 - 48x^3 - 24x + 48. \end{gathered}$$

As a final example take $A = K[x]$ with K a field and $p = x$ (which we have seen is prime[27]). Then $\delta_p : a \mapsto a'$ is the standard derivation of Example (a), e.g., for $K = \mathbb{Z}/5\mathbb{Z}$ we have $(x^5)' = 5x^4 = 0$.

[25] The terminology is not standard.

[26] See the comments following the statement of Theorem 3.5.

[27] E.g., following the statement of Theorem 3.5.

Returning to generalities, note from (i) that

(iii) $$p' = 1,$$

and that when A is an integral domain with quotient field of characteristic[28] zero one has

(iv) $$a' \neq 0 \qquad \Leftrightarrow \qquad p|a.$$

Finally, note that when $p \in \mathbb{Z}^+$ is a prime and $\delta_p : a \in \mathbb{Z} \to a' \in \mathbb{Z}$ is the p-adic semi-derivative one has

(v) $$(p^p)' = p^p.$$

(c) Let $R[x]$ and $d/dx : R[x] \to R[x]$ (but not A) be as in Example (a) and let $2 \leq n \in \mathbb{Z}$. Let A be the ring of $n \times n$ matrices (a_{ij}) with entries in $R[x]$. Then a derivation is defined on the ring A by $(a_{ij}) \mapsto (\frac{d}{dx}a_{ij})$. This is the basic example of a derivation on a non-commutative ring.

(d) The zero mapping $a \in A \mapsto 0 \in A$ is a derivation on A, and therefore a semi-derivation on this ring. This is the *trivial derivation*, or *trivial semi-derivation*, on A. Any other derivation or semi-derivation on A is *non-trivial*.

(e) The sum of two semi-derivations on A is again a semi-derivation on A, as is the product of any scalar with any semi-derivation. In particular, the semi-derivations on A form an A-module.

(f) The assertions of Item (e) also hold for derivations. Moreover, in that case the commutator or *bracket* $[\delta_1, \delta_2] := \delta_1 \circ \delta_2 - \delta_2 \circ \delta_1$ is also a derivation, and this binary operation $(\delta_1, \delta_2) \mapsto [\delta_1, \delta_2]$ of derivations δ_1 and δ_2 gives the collection of derivations on A the structure of an A-Lie algebra[29].

---

[28] The *characteristic* $\mathrm{char}(K)$ of a field K is defined to be 0 if $n \cdot 1 \neq 0$ for all $n \in \mathbb{Z}^+$; otherwise it is defined to be the smallest integer $n \in \mathbb{Z}^+$ such that $n \cdot 1 = 0$. It is not diffcult to prove that $\mathrm{char}(K)$ is either zero or a prime number, and we will assume this result. See, e.g., [Lang, Chapter II, §1, pp. 89-90].

[29] A left A-module $\mathfrak{L}$ is an *A-Lie algebra* if there is an A-bilinear mapping $(a, b) \in \mathfrak{L} \times \mathfrak{L} \mapsto [a, b] \in \mathfrak{L}$ satisfying

(a) $[\ell, m] = -[m, \ell]$ and

(b) $[\ell, [m, n]] + [n, [\ell, m]] + [m, [n, \ell]] = 0$

for all $\ell, m, n \in \mathfrak{L}$. Equality (a) is the *skew-symmetry* condition; (b) is the *Jacobi identity*. Example: $\mathbb{R}^3$ with the usual cross-product is an $\mathbb{R}$-Lie algebra.

The definition of the p-adic semi-derivation $\delta_p : A \to A$ given in Example 4.3(b) might appear to the reader to depend on the indicated set P of representatives of the primes. In fact this is not the case.

Proposition 4.4 : *Let A be a* UFD *and and let $p \in A$ be a prime. Then the p-adic value $\nu_p(a)$ and the p-adic semi-derivative δ_p are independent of the set P of representatives of the primes used in their definitions, so long as $p \in P$.*

The stated result for the p-adic valuation ν_p is inevitably assumed without comment in any text discussing "valuations"[30]. Once we have established the proof, we will adopt that custom.

Proof : Let P be as in Example 4.3(b) and let $Q \subset A$ be some other set of representatives of the primes which also contains p. If a non-zero non-unit $a \in A$ has factorizations $a = u \cdot \prod_{j=1}^{m} p_j^{n_j}$ with $p_j \in P$ and $a = v \cdot \prod_{i=1}^{\hat{m}} q_i^{\hat{n}_i}$ with $q_i \in Q$, then by unique factorization we must have $m = \hat{m}$ and, possibly after reordering, $p_j = u_j q_j$ with $u_j \in A^\times$ and $n_j = \hat{n}_j$ for $j = 1, 2, \ldots, m$.

If $p \not| a$ the prime p does not appear in either of the factorizations

$$\text{(i)} \qquad a = u \cdot \textstyle\prod_{j=1}^{m} p_j^{n_j} \qquad \text{and} \qquad a = v \cdot \prod_{j=1}^{m} q_j^{n_j},$$

hence $\nu_p(a) := 0$ and $\delta_p(a) := 0$ using either expression.

If $p|a$ then by a further reordering, if necessary, we can assume $p_m = p$ in (i). However, since p_m and q_m are associates, and since P and Q contain exactly one representative of each prime element, this forces $q_m = p$, hence $\nu_p(a) := n_m$ using either expression. We can now write (i) as $a = bp^{\nu_p(a)} = cp^{\nu_p(a)}$, whence $c = b$ (because A is an integral domain), and it is then evident from (i) of Example 4.3(b) that $\delta_p(a)$ is independent of the choice of P.

Finally, if a is a unit then $\delta_p(a) := 0$ when defined using either P or Q, and this completes the proof. **q.e.d.**

Corollary 4.5 : *Let A be a* UFD *with quotient field of characteristic zero and let $p, q \in A$ be primes. Then the following assertions are equivalent:*

(a) *p and q are associates;*

(b) *$\nu_p = \nu_q$; and*

[30]The p-adic valuations we have introduced are special cases of this more general concept. Since we have no need for the added generality we omit the definition.

(c) $\delta_p = u \cdot \delta_q$ *for some unit* $u \in A$.

For example, take $A = \mathbb{Z}$, $p = 3$ and $q = -3$. Then from $243 = 3^5$ we see that $\nu_3(243) = 5$ and that $\delta_3(243) = 5 \cdot 3^4 = 405$, whereas from $243 = (-1)(-3)^5$ we see that $\nu_{-3}(243) = 5$ and that $\delta_{-3}(243) = 5 \cdot (-1) \cdot (-3)^4 = -405 = -\delta_3(243)$.

Proof : Let $P \subset A$ be a set of representatives of the primes containing p.

(a) $\Rightarrow$ (b) : By assumption $p = uq$ for some unit $u \in A^\times$. Set $Q := (P \backslash \{p\}) \cup \{q\}$, which we note is a set of representatives of the primes containing q. Then an element $a \in A$ can be written (by means of P) in the form $a = bp^n$, where $p \not| b$, if and only if it can be written (by means of Q) in the form

$$\text{(i)} \qquad a = u^n b q^n = c q^n,$$

where $q \not| c$. Thus $\nu_p(a) = \nu_q(a)$.

(b) $\Rightarrow$ (c) : For $u \in A^\times$ as in (i) of the proof of (b) we see from

$$\delta_q a = n u^n b q^{n-1} = n u^n b (u^{-1} p)^{n-1} = n u^n u^{1-n} b p^{n-1} = u \cdot n b p^{n-1} = u \cdot \delta_p a$$

that $\delta_q = u \cdot \delta_p$.

(c) $\Rightarrow$ (a) : From (iv) of Example 4.3(b) we see that for any $a \in A$ we have $p|a$ if and only if $q|a$. Since $p|p$ this gives $q|p$ and, similarly, $p|q$. Assertion (a) follows. **q.e.d.**

The next result indicates how p-adic derivations might be applied to arithmetic problems.

Proposition 4.6 : *Let* A *be a* UFD *with quotient field of chacteristic zero, let* $P \subset A$ *be a set of representatives of the primes, and let* $B \subset A$. *Then:*

(a) *there is a non-zero non-unit common divisor of* B *if and only if there is a* $p \in P$ *such that* $\delta_p b \neq 0$ *for all* $b \in B$; *and*

(b) B *is relatively prime if and only if for each* $p \in P$ *there is an element* $b \in B$ *such that* $\delta_p b = 0$.

Proof : Use Proposition 2.14 and (iv) of Example 4.3(b). **q.e.d.**

We will give several more elementary number-theoretic applications of p-adic semi-derivatives following the next result.

Proposition 4.7 : *Any semi-derivation* $\delta : a \in A \mapsto a' \in A$ *has the following properties.*

(a) $0' = 0$.

(b) $1' = 0$.

(c) *When* $a \in A$ *is a unit one has*

(i) $$(u^{-1})' = -u^{-1} \cdot u' \cdot u^{-1}.$$

In particular, when A *is commutative one has the "reciprocal rule"*

(ii) $$(u^{-1})' = -u' \cdot u^{-2}.$$

(d) *For any* $a \in A$ *and any positive integer* n *one has*

(iii) $$(a^n)' = \sum_{j=0}^{n-1} a^{n-1-j} \cdot a' \cdot a^j.$$

In particular, when A *is commutative one has the "chain* (or "power") *rule"*

(iv) $$(a^n)' = n a^{n-1} a'.$$

(e) *For any positive integer* n *and any elements* $a_1, a_2, \ldots, a_n \in A$ *one has*

(v) $$(\prod_{j=1}^n a_j)' = \sum_{i=1}^n a_1 a_2 \cdots a_{i-1} a_i' a_{i+1} \cdots a_n.$$

In particular, when A *is commutative one has*

(vi) $$(\prod_{j=1}^n a_j)' = \sum_{i=1}^n (\prod_{j \neq i} a_j) a_i'.$$

Proof :

(a) From $0 = 0 \cdot 0$ we have $0' = 0 \cdot 0' + 0' \cdot 0 = 0 + 0 = 0$.

(b) From $1 = 1 \cdot 1$ we have $1' = 1 \cdot 1' + 1' \cdot 1 = 1' + 1'$, and (b) follows.

(c) From $1 = u \cdot u^{-1}$ and (b) we have $0 = u \cdot (u^{-1})' + u' \cdot u^{-1}$.

(d) For $n = 1$ formula (iii) reduces to the tautology $a' = a'$. If $n \geq 1$ and the formula holds for n we have

$$\begin{aligned}
(a^{n+1})' &= (a^n \cdot a)' \\
&= a^n \cdot a' + (a^n)' \cdot a \\
&= a^n \cdot a' + \left(\sum_{j=0}^{n-1} a^{n-1-j} \cdot a' \cdot a^j \right) \cdot a \\
&= a^n \cdot a' + \sum_{j=0}^{n-1} a^{n-1-j} \cdot a' \cdot a^{j+1} \\
&= a^n \cdot a' + \sum_{j=1}^{n} a^{n-j} \cdot a' \cdot a^j \\
&= \sum_{j=0}^{n} a^{n-j} \cdot a' \cdot a^j,
\end{aligned}$$

precisely as asserted.

(e) Use induction on n.

q.e.d.

Here is an amusing consequence of the chain-rule.

Corollary 4.8 : *Suppose A is a* UFD *with quotient field of characteristic zero, $a, p \in A$, p is a prime, and $p|a^n$ for some positive integer n. Then $p|a$.*

Proof : By (iv) of Example 4.3(b) (with a in that formula replaced by a^n) the hypothesis $p|a^n$ is equivalent to $\delta_p(a^n) \neq 0$. From the chain-rule we therefore have $0 \neq \delta_p(a^n) = na^{n-1}\delta_p a$, hence $\delta_p a \neq 0$, and with a second appeal to (iv) of Example 4.3(b) (in this instance as stated) we conclude that $p|a$. **q.e.d.**

Corollary 4.9 : *Suppose A is a* UFD *and $\delta : a \in A \mapsto a' \in A$ is a non-trivial semi-derivation. Then there is either a unit $u \in A$ such that $u' \neq 0$ or a prime $p \in A$ such that $p' \neq 0$,*

Non-trivial semi-derivations are defined in Example 4.3(d).

The two possibilities given in the statement are not mutually exclusive.

Proof : By assumption there is an element $a \in A$ such that $a' \neq 0$, and $a \neq 0$ by Proposition 4.7(a). If a is a unit we are done; otherwise we can factor a as in (i) of Proposition 2.13 and the result is then seen from (v) of Proposition 4.7(e). **q.e.d.**

Corollary 4.10 : *The only semi-derivation on a finite field, and hence the only derivation on such a field, is the trivial derivation.*

Proof : When A is a finite field of characteristic p the Frobenius mapping $a \mapsto a^p$ is a field isomorphism, and as a result any $a \in A$ can be written in the form $a = b^p$. For any semi-derivation $\delta : A \to A$ we then see from (iv) of Proposition 4.7(d) that $a' = pb^{p-1}b' = 0 \cdot b' = 0$. **q.e.d.**

Suppose A is an integral domain with quotient field K and $\delta : A \to A$ is a semi-derivation (hence possibly a derivation). The associated *logarithmic semi-derivative* is the mapping $\ell\delta : A \setminus \{0\} \to K$ defined by

$$\ell\delta : a \in A \setminus \{0\} \mapsto \delta(a)/a \in K. \tag{4.11}$$

When δ is a derivation this is the associated *logarithmic derivative*. As one might expect from the name, one has

$$\ell\delta(ab) = \ell\delta(a) + \ell\delta(b), \qquad a, b, \in A \setminus \{0\}, \tag{4.12}$$

as one sees from

$$\begin{aligned} \ell\delta(ab) &= \delta(ab)/ab \\ &= (a \cdot \delta b + \delta a \cdot b)/ab \\ &= \delta b/b + \delta a/b \\ &= \delta a/a + \delta b/b \\ &= \ell\delta(a) + \ell\delta(b). \end{aligned}$$

Proposition 4.13 : *Let $p \in \mathbb{Z}^+$ be a prime and let $\delta_p : \mathbb{Z} \to \mathbb{Z}$ be the p-adic semi-derivation. Then for any $0 \neq a \in \mathbb{Z}$ one has*

$$\ell\delta_p(a) = \nu_p(a)/p.$$

Proof : Immediate from (i) of Examples 4.3(b). **q.e.d.**

Corollary 4.14 : *Suppose $p \in \mathbb{Z}^+$ is prime and $n, k \in \mathbb{Z}^+$ are such that $k \nmid \nu_p(n)$. Then n is not the k^{th}-power of any element of $\mathbb{Q}$.*

To see an application take $p = n = 2$. We have $\nu_2(n) = 1$, and the hypotheses are therefore satisfied by any integer $k > 1$. We conclude that 2 is not the k^{th} power of any element of $\mathbb{Q}$.

Proof : Otherwise there are integers r, s such that $ns^k = r^k$. Applying the logarithmic derivative $\ell\delta_p$ gives

$$\nu_p(n)/p + k\nu_p(s)/p = k\nu_p(r)/p,$$

and it follows that $\nu_p(n) = k(\nu_p(r) - \nu_p(s))$. Since $\nu_p(r), \nu_p(s) \in \mathbb{Z}$, this contradicts the hypothesis $k \nmid \nu_p(n)$. **q.e.d.**

5. Extensions to Rings of Fractions

In this section A denotes a commutative ring with unity.

Proposition 5.1 : *Suppose $S \subset A\backslash\{0\}$ is a multiplicative subset and $\varphi : A \to S^{-1}A$ is the canonical homomorphism. Then to any semi-derivation $\delta : a \in A \to a' \in A$ there corresponds a unique semi-derivation $\delta_S : S^{-1}A \to S^{-1}A$ which renders the diagram*

$$\text{(i)} \qquad \begin{array}{ccc} S^{-1}A & \xrightarrow{\delta_S} & S^{-1}A \\ {\scriptstyle\varphi}\uparrow & & \uparrow{\scriptstyle\varphi} \\ A & \xrightarrow{\delta} & A \end{array}$$

commutative. Specifically, $\delta_S : S^{-1}A \to S^{-1}A$ is well-defined by the "quotient rule"

$$\text{(ii)} \qquad \delta_S : a/b \mapsto (ba' - b'a)/b^2, \qquad (a,b) \in A \times S.$$

When δ is a derivation the same is true of δ_S.

When φ is an embedding one expresses diagram (i) as

$$\text{(5.2)} \qquad \begin{array}{ccc} S^{-1}A & \xrightarrow{\delta_S} & S^{-1}A \\ {\scriptstyle\text{inc}}\uparrow & & \uparrow{\scriptstyle\text{inc}} \\ A & \xrightarrow{\delta} & A \end{array}$$

the idea being to regard the semi-derivation δ_S as an extension of δ. To keep in mind this important special case one refers to δ_S the *extension of δ by φ* even when φ is not an embedding. It is common practice to write δ_S as δ, and we will follow this custom.

Proof of Proposition 5.1 : To see that a mapping $\delta_S : S^{-1}A \to S^{-1}A$ is well-defined by (ii) recall that in $S^{-1}A$ we have $a/b = c/d$ if and only if

$$\text{(iii)} \qquad sad = sbc \qquad \text{for some} \qquad s \in S.$$

Multiplying by s then gives $s^2ad = s^2bc$, whereupon applying δ yields

$$\begin{aligned}
& s^2(ad)' + 2s'sad &=& \; s^2(bc)' + 2s'sbc \\
\Rightarrow \quad & s^2(ad)' &=& \; s^2(bc)' \;\; \text{(by (iii))} \\
\Rightarrow \quad & s^2bd(ad)' &=& \; s^2bd(bc)' \\
\Rightarrow \quad & s^2bd(ad' + a'd) &=& \; s^2bd(bc' + b'c) \\
\Rightarrow \quad & sad \cdot sbd' + s^2d^2 \cdot ba' &=& \; s^2bd \cdot bc' + sbc \cdot sb'd \\
\Rightarrow \quad & sbc \cdot sbd' + s^2d^2 \cdot ba' &=& \; s^2bd \cdot bc' + sad \cdot sb'd \;\; \text{(by (iii))} \\
\Rightarrow \quad & s^2b^2 \cdot d'c + s^2d^2 \cdot ba' &=& \; s^2b^2 \cdot dc' + s^2d^2 \cdot b'a \\
\Rightarrow \quad & s^2[d^2(ba' - b'a)] &=& \; s^2[b^2(dc' - d'c)].
\end{aligned}$$

Since S is multiplicative we have $s^2 \in S$, from the last equality we may therefore conclude

$$(ba' - b'a)/b^2 = (dc' - d'c)/d^2 \in S^{-1}A,$$

and in this way we see that δ_S is well-defined.

To verify the Leibniz rule choose any a/b and $c/d \in S^{-1}A$ and to ease notation write $\delta_S : (a/b)$ as $(a/b)'$, etc. Then

$$\begin{aligned}
((a/b) \cdot (c/d))' &= (ac/bd)' \\
&= (bd(ac)' - (bd)'ac)/(bd)^2 \\
&= (bd(ac' + a'c) - (bd' + b'd)ac)\,/(bd)^2 \\
&= (ab(dc' - d'c) + (ba' - b'a)cd)\,/(bd)^2 \\
&= (a/b) \cdot ((dc' - d'c)/d^2) + ((ba' - b'a)/b^2) \cdot (c/d) \\
&= (a/b) \cdot (c/d)' + (a/b)' \cdot (c/d),
\end{aligned}$$

precisely as desired.

To verify the commutativity of diagram (i) simply note that for any $a \in A$ one has

$$(\varphi(a))' = (a/1)' = (1 \cdot a' - 1' \cdot a)/1^2 = a'/1 = \varphi(a'),$$

and $\delta_S \circ \varphi = \varphi \circ \delta$ follows.

To establish uniqueness suppose diagram (i) commutes when δ_S is replaced by

a derivation $\hat{\delta} : S^{-1}A \to S^{-1}A$. Then for any $(a, b) \in A \times S$ we have

$$\begin{aligned}
\hat{\delta}(a/b) &= \hat{\delta}((a/1) \cdot (1/b)) \\
&= (a/1) \cdot \hat{\delta}(1/b) + \hat{\delta}(a/1) \cdot (1/b) \\
&= \varphi(a) \cdot (-\hat{\delta}(b/1) \cdot (b/1)^{-2}) + \hat{\delta}(\varphi(a)) \cdot (1/b) \\
&\qquad \text{(by (ii) of Proposition 4.7(c))} \\
&= -\varphi(a) \cdot (\hat{\delta} \circ \varphi)(b) \cdot (b/1)^{-2} + (\hat{\delta} \circ \varphi)(a) \cdot (1/b) \\
&= -\varphi(a) \cdot (\varphi \circ \delta)(b) \cdot (b/1)^{-2} + (\varphi \circ \delta)(a) \cdot (1/b) \\
&= -\varphi(a) \cdot \varphi(b') \cdot (b/1)^{-2} + \varphi(a') \cdot (1/b) \\
&= -\varphi(ab') \cdot (b/1)^{-2} + a'/b \\
&= -ab'/b^2 + a'/b \\
&= (ba' - b'a)/b^2 \\
&= \delta_S(a/b),
\end{aligned}$$

and uniqueness is thereby established.

To complete the proof suppose δ is a derivation, i.e., that the mapping $\delta : A \to A$ is an additive group homomorphism. Then for any a/b, $c/d \in S^{-1}A$ we have

$$\begin{aligned}
(a/b + c/d)' &= ((ad + bc)/bd))' \\
&= \big(bd(ad + bc)' - (bd)'(ad + bc)\big)/(bd)^2 \\
&= (bd(ad' + a'd + bc' + b'c) - (bd' + b'd)(ad + bc))\,/(bd)^2 \\
&= (abdd' + bd^2a' + b^2dc' + bcdb' - abdd' - b^2cd' - ad^2b' - bcdb')\,/(bd)^2 \\
&= (bd^2a' + b^2dc' - b^2cd' - ad^2b')\,/(bd)^2 \\
&= (d^2(ba' - b'a) + b^2(dc' - d'c))\,/(bd)^2 \\
&= (ba' - b'a)/b^2 + (dc' - d'c)/d^2 \\
&= (a/b)' + (c/d)',
\end{aligned}$$

and we conclude that $\delta_S : S^{-1}A \to S^{-1}A$ is also additive. **q.e.d.**

Corollary 5.3 : *When A is an integral domain any (semi-) derivation $\delta : a \in A \to a' \in A$ extends uniquely to a (semi-) derivation of the quotient field K of A, and this extension is well-defined by the "quotient rule"*

(i) $$a/b \mapsto (ba' - b'a)/b^2, \qquad a, b \in A, \qquad b \neq 0.$$

Examples 5.4 :

(a) Let R be an integral domain, let $A := R[x]$ (one indeterminate), and let $\delta = \frac{d}{dx}$ be the usual derivation. Then δ extends, via the quotient rule, to a derivation on the quotient field $R(x)$, and is the only derivation on $R(x)$ extending $\frac{d}{dx}$. This extension is also denoted $\frac{d}{dx}$, and is again called the *usual derivation.*

(b) Let $p \in \mathbb{Z}^+$ be a prime number and let $\delta_p : \mathbb{Z} \to \mathbb{Z}$ be the associated p-adic semi-derivation. Then the unique extension of δ_p to $\mathbb{Q}$ is given by $r = s/t \mapsto (t \cdot \delta_p s - \delta_p t \cdot s)/t^2$. Equivalently, write $r = (a/b)p^n$, where $a, b, p \in \mathbb{Z}$ are pairwise relatively prime, $b > 0$, and $n \in \mathbb{Z}$; then the extension of δ_p to $\mathbb{Q}$ is given by $r \mapsto n(a/b)p^{n-1}$. The extension again denoted δ_p, and is again called the *p-adic semi-derivation.*

6. Constants

A is a ring as in the previous section, i.e., commutative with unity.

The *constants* of a derivation or semi-derivation $\delta : a \in A \mapsto a' \in A$ are those elements of A in the kernel of δ, i.e., those $a \in A$ such that $a' = 0$. The totality of constants is denoted by A_C, or by A^δ when δ requires clarification. A necessary and sufficient condition for a derivation to be trivial is obviously $A_C = A$.

Examples 6.1 :

(a) When $R = \mathbb{Z}$ the constants of the derivation d/dx of Example 4.3(a) are the constant polynomials. However, for other choices of R the situation can be a bit more complicated. For example, when R is the ring $\mathbb{Z}/4\mathbb{Z}$ the "non-constant" polynomials $[2]x^2$ and x^4 are also constants, since $(d/dx)[2]x^2 = 2[2]x = [4]x = [0]$ and $(d/dx)x^4 = 4x^3 = [4]x^3 = [0]$.

(b) Suppose A is a UFD with quotient field of characteristic zero $p \in A$ is prime, and $a \mapsto a'$ is the p-adic derivation. Then a non-zero element a is a constant if and only if a is not divisible by p.

(c) A semi-derivation $\delta : A \to A$ is trivial if and only if all elements of A are constants.

Proposition 6.2 : *When $\delta : A \to A$ is a derivation any two primitives of an element of A differ by a constant.*

The result is false for semi-derivations, e.g., consider the 5-adic semi-derivation $\delta_5 : n \in \mathbb{Z} \to n' \in \mathbb{Z}$. One has $3' = 8' = 0$, whereas $(8-3)' = 5' = 1 \neq 0$.

Proof : If $b_1, b_2 \in A$ satisfy $b_1' = b_2'$ then from additivity we have $(b_1 - b_2)' = b_1' - b_2' = 0$. **q.e.d.**

Proposition 6.3 : *The collection A_C of constants of any semi-derivation is a multiplicative set i.e., it contains 1 and is closed under multiplication. Moreover, one always has $0 \in A_C$. If $a \in A_C$ is a unit then $a^{-1} \in A_C$.*

Proof : The initial assertion is immediate from Proposition 4.7(b) and the Leibniz rule (4.2); the second is a restatement of Proposition 4.7(b); the third is evident from (i) of Proposition 4.7(c). **q.e.d.**

Corollary 6.4 : *The constants* A_C *of any derivation on* A *form a subring of* A*, a sub-domain if* A *is an integral domain, and a subfield if* A *is a field. Moreover,* A_C *contains the image of the standard homomorphism* $n \in \mathbb{Z} \mapsto n \cdot 1 \in A$*.*

A_C is the *ring, domain* or *field of constants* in accordance with the assumed structure on A.

The final assertion of the corollary can be false for semi-derivations, e.g., $2 \notin \mathbb{Z}_C$ when $\delta : \mathbb{Z} \to \mathbb{Z}$ is the 2-adic semi-derivation.

Proof : When δ is a derivation the closure of A_C under addition is evident from (4.1); the inclusions $n \cdot 1 \in A_C$ are then seen from Proposition 4.7(b). Since any subring of and integral domain is an integral domain, the remaining assertions are immediate consequences of Proposition 6.3. **q.e.d**.

We next relate units with constants. This requires a few preliminaries.

Suppose A is commutative with unity $1 \neq 0$. The *characteristic*[31] $\mathrm{char}(A)$ *of* A is defined to be 0 if[32] $n \cdot 1 \neq 0$ for all $n \in \mathbb{Z}^+$; otherwise it is defined as the smallest positive integer n such that $n \cdot 1 = 0$. Note from $1 \neq 0$ that $\mathrm{char}(A) = 1$ is impossible. Examples: $\mathrm{char}(\mathbb{Z}) = 0$; $\mathrm{char}(\mathbb{Z}/n\mathbb{Z}) = n$.

Proposition 6.5 : *When* A *is commutative with unity* $1 \neq 0$ *the following assertions hold.*

(a) $\mathrm{char}(A) = 0$ *if and only if there is a non-zero element* $a \in A$ *such that* $n \cdot a \neq 0$ *for all* $n \in \mathbb{Z}^+$*.*

(b) $\mathrm{char}(A) = n > 0$ *if and only if* n *is the smallest positive integer such that* $n \cdot a = 0$ *for all* $a \in A$*. In particular,* $n \cdot a = 0$ *holds for all* $a \in A$*.*

(c) *Suppose* A *is an integral domain,* $n \in \mathbb{Z}^+$*, and there is a non-zero element* $a \in A$ *satisfying both* $n \cdot a = 0$ *and* $m \cdot a \neq 0$ *for any* $m \in \mathbb{Z}^+$ *with* $m < n$*. Then* $\mathrm{char}(A) = n$*.*

(d) *Suppose* A *is an integral domain,* $\mathrm{char}(A) = 0$*,* $a \in A$*, and* $n \in \mathbb{Z}^+$*. Then* $n \cdot a = 0$ *implies* $a = 0$*.*

(e) *Suppose* A *is an integral domain,* $\mathrm{char}(A) = n > 0$*,* $a \in A$*, and* $m \in \mathbb{Z}^+$ *satisfies* $m < n$*. Then* $m \cdot a = 0$ *implies* $a = 0$*.*

[31] We have already encountered this concept when A is a field: see Footnote 28. In this more general formulation we are following [Fral, Part IV, §19, pp. 181-2].

[32] See (ii) of Footnote 4 for the definition of $n \cdot 1$.

Proof : First note that for any $n \in \mathbb{Z}^+$ and any $a \in A$ we have

(i) $$n \cdot a = n \cdot (1a) = (n \cdot 1)a.$$

(a) $\Rightarrow$ Take $a = 1$.

$\Leftarrow$ From (i) we see that $n \cdot a \neq 0 \Rightarrow n \cdot 1 \neq 0$.

(b) $\Rightarrow$ From (i) we see that $n \cdot a = 0$ for all $a \in A$, and n is minimal w.r.t. this property since it is minimal when $a = 1$.

$\Leftarrow$ Take $a = 1$.

(c) By (i) and the integral domain hypothesis we have $n \cdot 1 = 0$, and by (i) with m replacing n we have $m \cdot 1 \neq 0$ for $1 \leq m < n$.

(d) $n \cdot 1 \neq 0$ in (i).

(e) Replacing n by m in (i) gives $0 = (m \cdot 1)a$, which by the integral domain hypothesis and $m \cdot 1 \neq 0$ forces $a = 0$.

q.e.d.

Corollary 6.6 : *Suppose A is an integral domain and $\delta : a \in A \mapsto a' \in A$ is a semi-derivation. Then the following assertions hold.*

(a) $-1 \in A_C$, *i.e.,* $(-1)' = 0$.

(b) *Suppose* $\mathrm{char}(A) = 0$. *Then any unit of finite order* (in the multiplicative group $A^\times$) *is a constant. In particular, when $A^\times$ is a finite group we must have $A^\times \subset A_C$.*

Proof :

(a) By (a) and (d) of Proposition 4.7 we have $0 = 1' = ((-1)^2)' = 2 \cdot (-1) \cdot (-1)'$, whereupon multiplication by -1 gives

$$0 = 2 \cdot (-1)' := (-1)' + (-1)'.$$

If $\mathrm{char}(A) = 0$ then $(-1)' = 0$ by Proposition 6.5(d); if $\mathrm{char}(A) = 2$ then from $1 + 1 = 0$ we see that $-1 = 1$, and (a) therefore holds by Proposition 4.7(b); if $\mathrm{char}(A) \geq 3$ the equality $(-1)' = 0$ follows from Proposition 6.5(e).

(b) When $u \in A^\times$ has order n we see by applying δ to $u^n = 1$ that

$$nu^{n-1}u' = 0.$$

From Proposition 6.5(d) this gives $u^{n-1}u' = 0$, and since $u^{n-1} \neq 0$ (because $u^{n-1} \in A^\times$) it follows that $u' = 0$.

q.e.d.

7. Semi-Derivations with Finite Support

Here A is a UFD*, P is a set of representatives of the primes of A, and $\delta : a \in A \mapsto a' \in A$ is a semi-derivation.*

In this brief section we are concerned with the following question: what role do the p-adic semi-derivations play within the collection of all semi-derivations on A? Our answer is given by Theorem 7.2.

The *P-support* of δ is the collection of all $p \in P$ such that $p' \neq 0$. We say that δ has *finite P-support* if

- the P-support of δ is finite, and
- $A^\times \subset A_C$, i.e., all units of A are constants.

The trivial semi-derivation is understood to have finite P-support.

Examples 7.1 : Let $P \subset \mathbb{Z}^+$ be the collection of positive primes.

(a) Any p-adic semi-derivation $\delta_p : \mathbb{Z} \to \mathbb{Z}$ has finite P-support.

(b) The semi-derivation $\delta := \sum_{p \in P} \delta_p$ does not have finite P-support. (This infinite sum is well-defined as a mapping of $\mathbb{Z}$ into $\mathbb{Z}$ since the number of primes dividing any fixed integer n is finite, hence $\delta_p n = 0$ for all but at most finitely many $p \in P$. For example, $\delta(75) = \delta(3{\cdot}5^2) = \delta_3 3{\cdot}5^2 + \delta_5 3{\cdot}5^2 = 5^2 + 2{\cdot}3{\cdot}5 = 55$.)

Theorem 7.2 : *The collection of semi-derivations on A with finite P-support is a free A-module having the p-adic semi-derivations $\{ \delta_p : p \in P \}$ as a basis.*

Proof : Choose any finite subset $\{\delta_{p_1}, \delta_{p_2}, \ldots, \delta_{p_n}\} \subset \{ \delta_p : p \in P \}$ and suppose $a_1, a_2, \ldots, a_n \in A$ are such that $\sum_{j=1}^n a_j \delta_{p_j} = 0$. Then for any $1 \leq i \leq n$ we have $0 = (\sum_{j=1}^n a_j \delta_{p_j}) p_i = a_i$, and we conclude that $\{ \delta_p : p \in P \}$ is linearly independent over A.

Let $\delta : a \in A \to a' \in A$ be a non-trivial semi-derivation with finite support and let $S_\delta = \{p_1, p_2, \ldots, p_n\}$ be that P-support. If $n = 1$ the semi-derivation $\delta - p_1' \delta_{p_1}$ is seen from Proposition 2.13 to be trivial, hence $\delta = p_1' \delta_{p_1}$. If $n > 1$ the difference $\delta - p_1' \delta_{p_1}$ is a semi-derivation of finite P-support, which by induction must be an A-linear combination of $\delta_{p_1}, \delta_{p_2}, \ldots, \delta_{p_{n-1}}$ (because the P-support is $\{p_1, p_2, \ldots, p_{n-1}\}$), and δ is thereby an A-linear combination of $\delta_{p_1}, \delta_{p_2}, \ldots, \delta_{p_n}$. **q.e.d.**

8. Prime Semi-Derivations

In this section A denotes an integral domain.

A semi-derivation $\delta : A \to A$ on A is *prime* if

- the complement in A of the collection of constants becomes an ideal when 0 is adjoined, and
- $1 \in A$ admits a primitive.

An example of a prime semi-derivation is provided by p-adic derivation $\delta_p : \mathbb{Z} \to \mathbb{Z}$ for any prime $p \in \mathbb{Z}^+$. The constants are those elements not divisible by p, and the complement, when 0 is adjoined, is the prime ideal (p). For an example in which the complement of the constants is not an ideal consider the semi-derivation $\delta_2 + \delta_3 : \mathbb{Z} \to \mathbb{Z}$: the elements 2 and 3 are not constants, but their non-zero sum does have this property.

The assumption that $1 \in A$ admit a primitive is a convenient normalization. To see that it is not automatic begin with any non-trivial semi-derivation $\delta : \mathbb{Z} \to \mathbb{Z}$ and note that all derivatives of the semi-derivation $2\delta : \mathbb{Z} \to \mathbb{Z}$ must be even.

The significance of the word "prime" in the definition of a prime semi-derivation is explained by the following result.

Proposition 8.1 : *The ideal occurring in the definition of a prime semi-derivation is a prime ideal.*

The ideal is said to be the *prime ideal associated* with δ, or simply the *associated prime ideal* when δ is clear from context.

Proof : Denote the ideal by $\mathfrak{i}$, and suppose $a, b \in A \setminus \{0\}$ satisfy $ab \in \mathfrak{i}$, i.e., suppose $ab = 0$ or $(ab)' \neq 0$. In the first case at least one of a and b must be 0, and therefore in $\mathfrak{i}$, by the integral domain assumption; in the second at least one of a' and b' must be non-zero, and therefore in $\mathfrak{i}$, by the Leibniz rule (4.2). **q.e.d.**

A semi-derivation $\delta : a \in \mathbb{Z} \to a' \in \mathbb{Z}$ is *non-negative* if $a > 0$ implies $a' \geq 0$. Any p-adic semi-derivation δ_p on $\mathbb{Z}$ provides an example, and the negation $-\delta_p$ provides a non-example.

Theorem 8.2 : *Let $P \subset \mathbb{Z}^+$ denote the collection of positive primes. Then the mapping $p \in P \mapsto (\delta_p : \mathbb{Z} \to \mathbb{Z})$ is a bijection between P and the non-negative prime semi-derivations on $\mathbb{Z}$.*

One should view this result in the same spirit as Theorem 3.6: one can generalize prime numbers by means of non-negative prime semi-derivations.

Proof : The mapping is clearly injective; what requires proof is surjectivity. To this end assume $\delta : a \in \mathbb{Z} \to a' \in \mathbb{Z}$ is a non-negative prime semi-derivation and let $(p) \subset \mathbb{Z}$ be the associated prime ideal, where w.l.o.g. $p \in \mathbb{Z}^+$. We will prove that δ is the p-adic semi-derivation δ_p.

Choose the minimal integer $0 < q \in (p)$ such that $q' = 1$. We claim that $q = p$. To verify this note from $q \in (p)$ that we can write q in the form $q = np^k$, where $0 \neq n \in \mathbb{Z}^+$ and $1 \leq k \in \mathbb{Z}^+$. (We do not require uniqueness here, e.g., we can allow that $p|n$.) Applying δ to q then gives

$$1 = \delta q = nkp^{k-1}p' + n'p^k. \tag{i}$$

It follows immediately that $k = 1$ (otherwise $p|1$), and equality (i) is thereby reduced to

$$1 = np' + n'p. \tag{ii}$$

Since n, p', n' and p are non-negative integers the same must be true of np' and $n'p$. It is then evident from (ii) and $p > 1$ that $n' = 0$, whence from $1 = np'$ and $p' \geq 0$ that $n = 1$ and $p' = 1$. The claim $q = p$ follows, and as a consequence we see that

$$p' = \delta_p p = 1.$$

Choose any $r \in (p)$ and write $r = mp^\ell$ with $\ell \geq 1$ and $p \not| \, m$. (Here we have uniqueness.) Then $m \notin (p)$, hence $m' = 0$, and

$$r' = m\ell p^{\ell-1}p' + m'p^\ell = m\ell p^{\ell-1} = \delta(mp^\ell) = \delta_p r$$

follows. If $r \notin (p)$ then $r' = 0$ (by the definition of a prime semi-derivation), and $\delta r = \delta_p r$ again follows. The asserted equality $\delta = \delta_p$ is thereby established. **q.e.d.**

9. Fermat's Little Theorem

Applications of derivations, in the form of derivatives, are certainly familiar to readers. Applications of semi-derivations might be another matter. As is illustrated by the proof of Corollary 4.14, they can be useful in formulating arguments depending on unique factorization. The proof of the following result is in the same spirit.

Theorem 9.1 (Fermat's Little Theorem) : *Let $p \in \mathbb{Z}^+$ be a prime and suppose $a \in \mathbb{Z}$ is not divisible by p. Then in $\mathbb{Z}/p\mathbb{Z}$ one has $[a]^{p-1} = [1]$.*

Fermat's Little Theorem is often used to test the primality of a given positive integer p: if for some integer $1 < a < p$ one has $[a]^{p-1} \neq [1]$ then p is not a prime number.

Proof : Assuming $a \in \mathbb{Z} \mapsto a' \in \mathbb{Z}$ is the p-adic semi-derivation, the non-divisibility hypothesis on a is equivalent to the assertion that $a \in \mathbb{Z}_C$, i.e., that $a' = 0$ (recall (iv) of Example 4.3(b)). Since $\mathbb{Z}_C$ is multiplicative (Proposition 6.3) it follows that all powers of a are also constants (which of course is also also evident from unique factorization), and as a result we have $[a]^k \neq [0]$ in $\mathbb{Z}/p\mathbb{Z}$ for all $k \geq 0$. By the pigeonhole principle[33] there must be integers $1 \leq i < j \leq p$ such that $[a]^i = [a]^j$, i.e., such that $a^j - a^i$ is divisible by p or, equivalently, such that $(a^j - a^i)' \neq 0$. Since $a^j - a^i = a^i(a^{j-i} - 1)$ we conclude from $a^i \in \mathbb{Z}_C$ and the Leibniz rule that $(a^{j-i} - 1)' \neq 0$, i.e., that $[a]^{j-i} = [1]$ in $\mathbb{Z}/p\mathbb{Z}$. Since the (multiplicative) group of units $(\mathbb{Z}/p\mathbb{Z})^\times$ of $\mathbb{Z}/p\mathbb{Z}$ is of order $p-1$ we now see from Lagrange's Theorem that $(j-i)|(p-1)$, say $p - 1 = (j-i)m$, and $[a]^{p-1} = ([a]^{j-i})^m = [1]^m = [1]$ follows. **q.e.d.**

Thus far we have been able to proceed assuming minimal algebraic prerequisites on the part of readers. In the following result and section we abandon that stance: we need to assume familiarity with algebraic closures of fields.

Corollary 9.2 : *For any odd prime p the following assertions hold.*

(a) *The collection $\{[1], [2], \ldots, [p-1]\} \subset \mathbb{Z}/p\mathbb{Z}$ is a complete set of solutions of the equation $x^{p-1} = [1]$, i.e., the polynomial $x^{p-1} - [1]$ has no additional roots in any algebraic closure of $\mathbb{Z}/p\mathbb{Z}$.*

[33] I.e., if you distribute more than n letters into n mailboxes at least one mailbox will end up containing more than one letter. The result can be stated formally in terms of the non-existence of injective functions in special circumstances, but the pigeonhole formulation seems much easier to understand.

(b) *For any integer* a *not divisible by* p *one has*

(i) $$[a]^{\frac{p-1}{2}} = \pm[1] \in \mathbb{Z}/p\mathbb{Z}.$$

(c) *The mapping* $f : [a] \in (\mathbb{Z}/p\mathbb{Z})^\times \mapsto [a]^{\frac{p-1}{2}} \in \{[1], [-1]\}$ *is a surjective multiplicative group homomorphism with kernel of index* 2.

(d) *An element* $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$ *is the square of an element of* $(\mathbb{Z}/p\mathbb{Z})^\times$ *if and only if* $[a]$ *is a solution of the equation*

(ii) $$x^{\frac{p-1}{2}} = [1],$$

i.e., if and only if $[a] \in \ker f$, *where* $f : (\mathbb{Z}/p\mathbb{Z})^\times \to \{[1], [-1]\}$ *is as in* (c).

Elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ which are squares of elements of this group are called *quadratic residues modulo* p, or simply *quadratic residues* when p is clear from context. These entities play a fundamental role in number theory; we will see a hint of their relevance in Theorem 10.1.

Proof :

(a) Since $\mathbb{Z}/p\mathbb{Z}$ is a field the equation has no more than $p-1$ solutions, and the given elements are solutions by Fermat's Little Theorem.

(b) The equation $x^2 = [1]$ has exactly two solutions in $\mathbb{Z}/p\mathbb{Z}$, i.e., $x = [1]$ and $x = -[1]$. However, by Fermat's Little Theorem $[a]^{(p-1)/2}$ is a solution for any $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$, and the result follows.

(c) and (d) : The verification that f is a group homomorphism is straightforward. In establishing the remaining statements we follow [Serre, Chapter I, §3, p. 6].

Let K be any algebraic closure of the field $\mathbb{Z}/p\mathbb{Z}$, choose any non-zero $[a] \in \mathbb{Z}/p\mathbb{Z}$, and then a solution $k \in K$ of the equation $x^2 = [a]$. Since this equation has exactly two solutions, one being the negative of the other, $[a]$ is a square in $(\mathbb{Z}/p\mathbb{Z})^\times$ if and only if $k \in \mathbb{Z}/p\mathbb{Z}$. Now note from (b) that

$$k^{p-1} = (k^2)^{\frac{p-1}{2}} = [a]^{\frac{p-1}{2}} = \pm[1].$$

If $k \in \mathbb{Z}/p\mathbb{Z}$ then $k^{p-1} = [1]$ by Fermat's Little Theorem. Conversely, if $k^{p-1} = [1]$ then $k \in \mathbb{Z}/p\mathbb{Z}$ by (a). Thus $k \in \mathbb{Z}/p\mathbb{Z}$ if and only if $k \in \ker f$, and (d) follows.

To complete the proof note that equation (ii) has at most $(p-1)/2$ solutions in $\mathbb{Z}/p\mathbb{Z}$, all necessarily non-zero, whereas the field has $p - 1 > (p-1)/2$ non-zero

elements. The group homomorphism of (c) is therefore surjective. Since the image has two elements the same is true of the isomorphic factor group $(\mathbb{Z}/p\mathbb{Z})^+/\ker f$, and we conclude that the kernel has index 2.

q.e.d.

10. Sums of Two Squares

This section is optional, and does not make explicit use of either derivations or semi-derivations. It presents a substantial consequence of Fermat's Little Theorem which has motivated a great deal of number theoretic work over several centuries. The material is offered in part because, given our work to this point, it takes very little additional effort to do so.

The problem addressed is: can one give a procedure for deciding when a given positive integer can be expressed as the sum of squares of two integers? The answer is yes, and we will give a complete solution[34]. In asserting that an integer is the sum of two squares (of integers) we allow for the possibility that one of the other integers is 0, e.g., we view $4 = 2^2 + 0^2$ as expressing 4 as a sum of two squares.

We begin with a few observations.

I. From $2 = 1^2 + 1^2$ we see that the integer 2 can be written as the sum of two squares.

II. The equation $n = a^2 + b^2$ has integer solutions if and only if there are integer solutions a, b with $a \geq 0$ and $b \geq 0$. (Obvious.)

III. The integer 3 cannot be written as the sum of two squares. Indeed, if $3 = a^2+b^2$ then $|a| < 2$ and $|b| < 2$, and by II we can then assume $a, b \in \{0, 1\}$. However, for $a, b \in \{0, 1\}$ we see that $a^2 + b^2 \in \{0, 1, 2\}$, and the assertion is thereby verified.

IV. For any real numbers a, b, c, d one has

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2.$$

In particular, if a positive integer n factors as $n = n_1 n_2$, and if each of n_1 and n_2 can be written as the sum of two squares, then the same holds for n.

V. Any power of 2 can be written as the sum of two squares. This is immediate from I and IV.

VI. The square of any integer can be written as the sum of two squares. Indeed, $n^2 = n^2 + 0^2$.

[34]The problem is a variation of the classical geometric problem of Pythagorean triples: find all integer 3-tuples (a, b, c) such that $a^2 + b^2 = c^2$. This classical forerunner will not be addressed in these notes.

It seems evident from IV and V that to solve our problem we need to know which odd prime numbers $p \in \mathbb{Z}^+$ can be written as the sum of two squares. This is the situation we handle first.

Number theorists use the term "integer" in a far more inclusive way than other mathematicians. For example, the "integers" of the ring $\mathbb{Q}[i]$ consist of the *Gaussian integers*, i.e., the collection $\{\, a + ib \,:\, a, b \in \mathbb{Z} \,\}$. So as to distinguish the general concept from the context of $\mathbb{Z}$ they refer to elements of $\mathbb{Z}$ as *rational integers*, and to the prime numbers within $\mathbb{Z}^+$ as *rational primes*. We will follow this custom from this point on.

Theorem 10.1 : *When $p > 2$ is a rational prime the following statements are equivalent.*

(a) *p is not a prime in $\mathbb{Z}[i]$;*

(b) *p can be written as a sum of squares of two rational integers;*

(c) *$[p] = [1] \in \mathbb{Z}/4\mathbb{Z}$; and*

(d) *-1 is a quadratic residue modulo p.*

This is a highly non-trivial result. For example, the rational prime[35] $1,798,361,809$ satisfies (c), and as a consequence we can be guaranteed, without any additional work, that it can be expressed as a sum of two squares. Imagine the time it would take to establish this by searching for potential solutions. In contrast, the prime $1,798,361,819$ does not satisfy (c), and as a result cannot be written as a sum of two squares.

Assertion (c) would more commonly be expressed as $p \equiv 1 \pmod 4$. In plain English the condition is that p be one more than a multiple of 4.

Proof : In the proof $N : \mathbb{Z}[i] \to \mathbb{N}$ denotes the norm function $a + bi \mapsto a^2 + b^2$. The multiplicative property

$$\text{(i)} \qquad N(uv) = N(u)N(v), \qquad u, v \in \mathbb{Z}[i],$$

is essential, as are the following two properties:

(ii) an element $u \in \mathbb{Z}[i]$ is a unit if and only if $N(u) = \pm 1$; and

(iii) an element $q \in \mathbb{Z}[i]$ is prime if $N(q)$ is a rational prime.

[35] The primality of $1,798,361,809$ and $1,798,361,819$ was established by computer calculation.

(Recall (2.6(b)), Proposition 2.7(b), and Example 2.5(b).)

(a) $\Rightarrow$ (b) : When (a) holds we can write $p = (a+ib)(c+id)$, with neither factor a unit, and from (i) and (ii) we then see that

$$p^2 = (a^2+b^2)(c^2+d^2) \qquad \text{with} \qquad a^2+b^2 > 1 \qquad \text{and} \qquad c^c+d^2 > 1.$$

The equalities $a^2+b^2 = p = c^2+d^2$ follow, and this gives (b).

(b) $\Rightarrow$ (a) : p factors in $\mathbb{Z}[i]$ as $p = (a+ib)(a-ib)$ and (by (ii)) neither factor is a unit.

(b) $\Rightarrow$ (c) : Since p is odd, a and b must have opposite parity, so assume w.l.o.g. that $a = 2c$ and $b = 2d+1$ with $c, d \in \mathbb{Z}$. When then have

$$p = (2c)^2 + (2d+1)^2 = 4c^2 + 4d^2 + 4d + 1,$$

hence $[p] = [1] \in \mathbb{Z}/4\mathbb{Z}$.

(c) $\Rightarrow$ (d) : By assumption $p = 4m+1$ for some positive integer m, and we therefore have

$$\tfrac{p-1}{2} = 2m.$$

It then follows from Corollary 9.2(d) that we can choose a non-zero integer a such that $[a]^{2m} = [-1]$ in $\mathbb{Z}/p\mathbb{Z}$. To establish (d) rewrite this as $([a]^m)^2 = [-1]$.

(d) $\Rightarrow$ (b) : Here the argument is adapted from [N-Z-M, Chapter 2, §2.1, Lemma 2.13, pp. 54-5].

By assumption there is an integer s such that

$$[s]^2 = [-1] \in \mathbb{Z}/p\mathbb{Z}. \tag{iv}$$

Let n be the greatest rational integer less than $\sqrt{p}$ and consider all rational integer pairs (u, v) with $0 \leq u, v \leq n$. Since there are $(n+1)^2 > p$ such pairs and only p elements in $\mathbb{Z}/p\mathbb{Z}$ we conclude from the pigeonhole principal that there must be distinct pairs (u_1, v_1), (u_2, v_2) such that $[u_1 - sv_1] = [u_2 - sv_2] \in \mathbb{Z}/p\mathbb{Z}$. By defining $a := u_1 - v_1$ and $b := v_1 - v_2$ we can write this as $[a] = [s][b] \in \mathbb{Z}/p\mathbb{Z}$, from which we see that $[a]^2 = [s]^2[b]^2 = -[b]^2$, the last equality by (iv). This gives $p|(a^2+b^2)$, and since $0 < a^2+b^2 < p+p = 2p$ it follows that $a^2+b^2 = p$.

q.e.d.

Corollary 10.2 : *Suppose $2 < p \in \mathbb{Z}$ is prime and $a, b \in \mathbb{Z}$ are such that $p|(a^2+b^2)$, whereas p does not divide at least one of a and b. Then $[p] = [1] \in \mathbb{Z}/4\mathbb{Z}$. In particular, if $[p] = [3]$ then $p|a$ and $p|b$.*

Proof : Assume w.l.o.g. that $p \not| a$. Then $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$, and as a result $[a]$ must admit a multiplicative inverse.

By assumption we have

$$[b]^2 = -[a]^2 \in \mathbb{Z}/p\mathbb{Z},$$

whereupon multiplying by $[a]^{-2}$ yields

$$([a]^{-1}[b])^2 = -[1].$$

We conclude that -1 is a quadratic residue modulo p, and the corollary follows. **q.e.d.**

Corollary 10.3 : *When $p > 2$ is a rational prime the following statements are equivalent.*

(a) *p is also a prime in $\mathbb{Z}[i]$;*

(b) *p cannot be written as the sum of the squares of two rational integers;*

(c) *$[p] = [3] \in \mathbb{Z}/4\mathbb{Z}$; and*

(d) *-1 is a not quadratic residue modulo p.*

We are now in a position to formulate and prove the "sum of two squares" problem listed as Item III at the beginning of the chapter.

Application III : *Let n be a positive rational integer and express the unique prime factorization of n in the form*

$$n = 2^r \cdot \prod_{p \equiv 1 \bmod 4} p^s \cdot \prod_{q \equiv 3 \bmod 4} q^t. \tag{10.4}$$

Then n can be written as a sum of squares of two rational integers if and only if all the exponents t are even.

Examples: the rational integer $12,380,495,000 = 2^3 \cdot 5^4 \cdot 19^5$ cannot be written as a sum of two squares (because the exponent of 19 is odd), whereas $235,229,405,000 = 2^3 \cdot 5^4 \cdot 19^6$ can be so written.

Fermat claimed this result, but it is not known if he ever wrote down a proof. For history see [Ed, Chapter 1, §1.7, pp. 14-19]. Our proof is adapted from [N-Z-M, Chapter 2, §2.1, Theorem 2.15, pp. 55-6].

Proof :

$\Rightarrow$ By assumption we have $n = a^2 + b^2$ with $a, b \in \mathbb{Z}$. If q is a prime factor of n satisfying $q \equiv 3 \pmod 4$ then $q|a$ and $q|b$ by Corollary 10.2, and we therefore have $q^2|n$. This gives $t \geq 2$ and if $t = 2$ we are done. If not make the same argument for the rational integer identity $(a/q)^2 + (b/q)^2 = n/q^2$ to conclude that $t \geq 4$, etc. After finitely many such steps the procedure must stop, and we conclude that t must be even.

$\Leftarrow$ Use V and VI (at the beginning of the section) together with Theorem 10.1.

q.e.d.

11. Differential and Semi-Differential Ideals

In this section A *is a commutative ring with unity and* $\delta : a \in A \mapsto a' \in A$ *is a semi-derivation.*

An ideal $\mathfrak{i} \subset A$ is a *δ-ideal* if $a' \in \mathfrak{i}$ whenever $a \in \mathfrak{i}$. When δ is understood such an ideal is also called a *semi-differential ideal*. When δ is a derivation a semi-differential ideal is called a *differential ideal*.

The zero ideal and A are always semi-differential. Additional examples are most easily presented by first recording the following observation.

Proposition 11.1 : *For any element* $a \in A$ *the following assertions are equivalent:*

(a) *the principal ideal* (a) *is semi-differential;*

(b) $a' \in (a)$*; and*

(c) $a|a'$.

Proof :

(a) $\Rightarrow$ (b) : Immediate from the definition of a semi-differential ideal.

(b) $\Leftrightarrow$ (c) : For any $b \in A$ one has $b \in (a)$ if and only if $a|b$.

(c) $\Rightarrow$ (a) : By assumption there is an element $c \in A$ such that $a' = ac$. For any $ab \in (a)$ one therefore has

$$(ab)' = a \cdot b' + a' \cdot b = a \cdot b' + ac \cdot b = a(b' + cb) \in (a).$$

q.e.d.

Corollary 11.2 : *When* δ *is a derivation the following assertions are equivalent for any element* $a \in A$*:*

(a) *the principal ideal* (a) *is differential;*

(b) $a' \in (a)$*; and*

(c) $a|a'$.

Proof : By definition an ideal is differential if and only if it is semi-differential when δ is considered a semi-derivation. **q.e.d.**

Corollary 11.3 : *Suppose* A *is a* UFD *of characteristic zero,* $p \in A$ *is a prime, and* $a \in A \mapsto a' \in A$ *is the* p*-adic semi-derivation. Then a non-zero principal ideal* (a) *is semi-differential if and only if* $p|\nu_p(a)$.

Proof : Write $a = bp^{\nu_p(a)}$, where $p \not| \, b$. Recall from Example 6.1(b) that $b' = 0$ must hold, and we therefore have $a' = b' = 0$ if $\nu_p(a) = 0$, and $a' = b \cdot \nu_p(a) \cdot p^{\nu_p(a)-1} + b' \cdot p_p^{\nu}(a) = b\nu_p(a)p^{\nu_p(a)-1}$ otherwise. Thus

$$a|a' \quad \Leftrightarrow \quad \begin{cases} \nu_p(a) = 0 & \text{or} \\ \nu_p(a) \neq 0 & \text{and} \quad p|\nu_p(a) \end{cases} \quad \Leftrightarrow \quad p|\nu_p(a).$$

q.e.d.

Examples 11.4 :

(a) Choose any prime $p \in \mathbb{Z}^+$ and let $a \mapsto a'$ denote the corresponding p-adic semi-derivation on $\mathbb{Z}$. Then the corresponding semi-differential ideals of $\mathbb{Z}$ are those of the form (bp^k) in which $p \not| \, b$ and k is a non-negative integer divisible by p. In particular, (p^p) is a semi-differential ideal, whereas (p^k) for $k \in \mathbb{Z}^+$ satisfying $k < p$ does not have this property.

(b) *When the usual derivation* d/dx *is assumed the only differential ideals of* $\mathbb{Q}[x]$ *are the zero ideal and the full polynomial ring.* This can also be deduced from Corollary 11.3, but it is just as easy to establish the result directly. Since $\mathbb{Q}[x]$ is a PID any non-zero differential ideal must be of the form (p) for some non-zero polynomial $p \in \mathbb{Q}[x]$. Since d/dx lowers the degrees of non-constant polynomials the only way we can have $p' \in (p)$ is if $p' = 0$. This would mean p is a non-zero constant, and $(p) = \mathbb{Q}[x]$ follows.

Semi-differential and differential ideals arise from semi-differential and differential ring homomorphisms. Specifically, let B be a ring, let $\gamma : B \to B$ be a semi-derivation, and let $f : A \to B$ be a ring homomorphism. f is a *semi-differential homomorphism* if

$$f \circ \delta = \gamma \circ f, \tag{11.5}$$

and is a *differential homomorphism* if this is the case and both δ and γ are derivations.

Proposition 11.6 : *Suppose* $f : A \to B$ *as above is a semi-differential homomorphism. Then* $\ker f$ *is a semi-differential ideal. Moreover, if* δ *and* γ *are derivations and* f *is a differential homomorphism then* $\ker f$ *is a differential ideal.*

Proof : $\ker f$ is an ideal by standard algebra, and for $a \in \ker f$ we see from (11.5) that $\delta a \in \ker f$. **q.e.d.**

When $\delta : A \to A$ is a derivation and $\mathfrak{i} \subset A$ is a differential ideal a derivation is well-defined on the residue class ring $A/\mathfrak{i}$ by

$$[a]' := [a'], \qquad [a] \in A/\mathfrak{i}. \tag{11.7}$$

The canonical homomorphism $a \in A \mapsto [a] \in A/\mathfrak{i}$ then provides an example of a differential homomorphism and, simultaneously, a semi-differential homomorphism.

The construction of the previous paragraph can fail for semi-derivations which are not derivations, the problem being that $[a]'$ may not be well-defined by (11.7). Indeed, if $[a] = [b]$ in $A/\mathfrak{i}$ then $a - b \in \mathfrak{i}$, hence $(a-b)' \in \mathfrak{i}$. However, to ensure $[a'] = [b']$ we need $a' - b' \in \mathfrak{i}$, and without additivity this does not follow from $(a-b)' \in \mathfrak{i}$.

12. Differential and Semi-Differential Units, Primes, Irreducibles, and Associates

In this section A denotes an integral domain.

In the integral domain case one can formulate the concepts of unit, prime, irreducible and associate in terms of ideals. An element $a \in A$ is:

- a unit if and only if a is not contained in any prime ideal[36];
- prime if and only if $a \neq 0$ and (a) is a prime ideal (Theorem 3.5(a));
- irreducible if and only if (a) is a proper ideal which is maximal among all principal ideals (Theorem 3.5(b)); and
- an associate of an element $b \in A$ if and only if $(a) = (b)$ (Proposition 2.2(a)).

In this section we use these characterizations to formulate the analogous concepts in differential and semi-differential rings. The results, we will find, are somewhat unsatisfactory, with no obvious means for resolution.

Our first task is to define prime and maximal differential and semi-differential ideals.

Suppose $\delta : A \to A$ is a derivation and $\mathfrak{i} \subset A$ is a proper δ-ideal. Then:

- $\mathfrak{i}$ is a *prime δ-ideal* if $\mathfrak{i}$ is a prime ideal;
- $\mathfrak{i}$ is a *maximal δ-ideal* if there are no proper δ-ideals properly containing $\mathfrak{i}$.

When δ is understood prime and maximal δ-ideals are also called *prime differential ideals* and *maximal differential ideals* respectively. By replacing "differential" with "semi-differential" we obtain the corresponding definitions for semi-differential ideals.

As is illustrated in Example 12.1(a), a maximal semi-differential ideal need not be a maximal ideal in the usual sense. We will see in Example 12.1(b) that a maximal semi-differential ideal need not be a prime semi-differential ideal.

[36]This requires the result that any ideal in a commutative ring with unity is contained in a maximal ideal, and therefore in a prime ideal. See, e.g., [Hun, Chapter III, §2, Theorem 2.18, p. 128].

Examples 12.1 :

(a) *Assume the usual derivation d/dx on $\mathbb{Q}[x]$. Then (0) is both a prime differential ideal and a maximal differential ideal, but is not a maximal ideal* (in the usual sense). This is clear from Example 11.4(b).

(b) *Let $p \in \mathbb{Z}^+$ be prime and give $\mathbb{Z}$ the corresponding p-adic semi-derivation. Then the ideal (p^p) is a maximal semi-differential ideal which is not a prime semi-differential ideal.* Indeed, this particular ideal was already seen to be semi-differential in Example 11.4(a). If it were properly contained in some other semi-differential ideal that ideal would have the form (a), with $a|p^p$. But a could then be assumed of the form p^k, with $0 \leq k < p$. If $k = 0$ then $(a) = \mathbb{Z}$, which is not a prime ideal, whereas if $k > 0$ the ideal is not semi-differential by Example 11.4(a). Finally, since the ideal is not prime (e.g., by Theorem 3.5(a)) it cannot be a prime semi-differential ideal.

Again suppose $\delta : A \to A$ is a derivation. We define a *δ-unit* of A to be an element not contained in any prime δ-ideal of this ring. When δ is understood we speak of a *differential unit*. By replacing "derivation" with "semi-derivation" and "differential" with "semi-differential" we obtain the definition of a *semi-differential unit*.

Examples 12.2 :

(a) *Assuming the usual derivation d/dx, all non-zero elements of $\mathbb{Q}[x]$ are differential units.* Indeed, we see from Example 11.4(b) that $(0) \subset \mathbb{Q}[x]$ is the unique maximal differential ideal, hence no non-zero element is contained in a prime differential ideal. Note that the units of $\mathbb{Q}[x]$ are the non-zero constant polynomials; we therefore have examples of differential units which are not units.

(b) *Any unit of A is a semi-differential unit for any semi-derivation and a differential unit for any derivation.* Since a unit is not contained in any prime ideal it cannot be contained in any prime semi-differential ideal or in any prime differential ideal.

(c) One can see from the discussion in Example 11.4(a) that when $p \in \mathbb{Z}^+$ is a prime and the corresponding p-adic semi-derivation is assumed the non-unit p^p is a semi-differential unit.

There are many equivalent definitions of a unit in A, one being that $a \in A$ is not contained in any maximal ideal, but the analogues for the semi-differential case are not all equivalent. One must therefore make a choice. We chose the definition above so as to guarantee the following result and corollary.

Proposition 12.3 : *The δ-units of a semi-derivation $\delta : A \to A$ form a multiplicative subset $S \subset A$, and $\hat{S} := S \setminus \{0\}$ is also multiplicative. The same assertion holds for δ-units when δ is a derivation.*

Recall that in this section A is assumed an integral domain.

Proof : If $a, b \in A$ are δ-units and $\mathfrak{p} \subset A$ is a δ-prime ideal then $a, b \notin \mathfrak{p}$, and therefore $ab \notin \mathfrak{p}$ (simply by virtue of the fact that $\mathfrak{p}$ is prime). This gives the initial assertion; the final assertion is then immediate from the integral domain hypothesis. **q.e.d.**

Corollary 12.4 : *In the notation of* Proposition 12.3 *consider A as a subring of the ring of fractions $\hat{S}^{-1}A$. Then the semi-differential units of A are identified with units* (in the usual sense) *of $\hat{S}^{-1}A$. Moreover, the semi-derivation on A extends to a semi-derivation on $\hat{S}^{-1}A$, and all the units of $\hat{S}^{-1}A$ are semi-differential units w.r.t. this extension. The same assertion holds when the prefix "semi-" is dropped from semi-derivation and semi-differential.*

Proof : Recall Proposition 5.1. **q.e.d.**

When $\delta : A \to A$ is a derivation the intersection of any family of differential ideals of A is again a differential ideal. When the family consists of those differential ideals containing a particular subset $S \subset A$ this intersection is denoted $[S]$ and is known as the *differential ideal generated by* S. When $S = \{s_1, s_2, \ldots, s_n\}$ is finite the notation $[\{s_1, s_2, \ldots, s_n\}]$ is condensed to $[s_1, s_2, \ldots, s_n]$. In particular, when $S = \{s\}$ is a singleton one writes $[s]$ in place of $[\{s\}]$. Ideals of this last form will play the role of principal ideals.

By replacing all occurrences of "differential" with "semi-differential" in the previous paragraph we define the analogous concepts for semi-derivations. We will use the same notation in both cases.

Examples 12.5 : Let $\delta_3 : \mathbb{Z} \to \mathbb{Z}$ be the 3-adic semi-derivation. Then:

(a) $[3] = \mathbb{Z}$, since $3' = 1$ (and $3' \in [3]$);

(b) $[3^2] = \mathbb{Z}$, since $(3^2)' = 2 \cdot 3 = 6$, $6' = 2$, and $1 = 3^2 - 4 \cdot 2 \in [3^2]$;

(c) $[3^3] = (27)$ (recall Example 11.4(a)); and

(d) $[2] = (2)$ (because $2' = 0 \in (2)$).

Let $\delta : A \to A$ be a derivation. An element $a \in A$ is:

- δ-*prime* if the differential ideal $[a]$ is prime;
- δ-*irreducible* if the differential ideal $[a]$ is both proper and maximal among all differential ideals of the form $[b]$;
- a δ-*associate* of an element $b \in A$ if $[a] = [b]$.

When δ is understood one speaks of *differential primes, differential irreducibles*, and *differential associates* respectively. The analogous concepts for semi-derivations are obtained by replacing all occurrences of "differential" by "semi-differential."

Examples 12.6 : Assume the 3-adic semi-derivation on $\mathbb{Z}$.

(a) The integers 3, 3^2 and 3^3 are semi-differential units, as one can see from the discussion in Example 12.1(b). Moreover, 3 and 3^2 are semi-differential associates (even though $3^2 \not| 3$), whereas 3 and 3^3, and 3^2 and 3^3, are not.

(b) One can also see from from the discussion in Example 12.1(b) that 3^3 is a semi-differential irreducible, but is not a semi-differential prime. In particular, and in stark contrast to the situation in ordinary commutative algebra, a ring element can simultaneously be a semi-differential irreducible and a semi-differential unit.

(c) $[6] = (2) = [2]$ (because $6' = 2$ and (2) is a semi-differential prime ideal), hence 6 is a semi-differential prime and 6 and 2 are semi-differential associates.

(d) Any prime $q \in \mathbb{Z}^+$ other than 3 is both a semi-differential prime and a semi-differential irreducible.

(e) Formulating a notion of "unique factorization" becomes problematic. For example, $27 = 27$ is a factorization of 27 into semi-differential irreducibles, but there is no factorization into semi-differential primes since 27 is a semi-differential unit.

Proposition 12.3, Corollary 12.4 and the following result are perhaps the only satisfactory consequences of the formulations in this section.

Proposition 12.7 : *Suppose A is a* PID *and $\delta : A \to A$ is a semi-derivation. Then any δ-prime in A is δ-irreducible.*

To appreciate this result one should recall Theorem 2.2(i).

Proof : This is immediate from the fact that prime ideals in a PID are maximal ideals. **q.e.d**.

Acknowledgment

I would like to thank Ms. Nelly Choueiri for carefully reading the manuscript, pointing out many errors, and suggesting many corrections. Any remaining errors are my responsibility alone.

References

[Bui] A. Buium, *Differential Algebra and Diophantine Geometry*, Hermann, Paris, 1994.

[Ed] H.M. Edwards, *Fermat's Last Theorem*, GTM 50, Springer-Verlag, New York, 1977.

[El] N.D. Elkies, Pythagorean Triples and Hilbert's Theorem 90, *Amer. Math. Monthly*, **110**, (2003), 678.

[Fral] J.B. Fraleigh, *A First Course in Abstract Algebra*, Seventh Edition, Addison-Wesley, Boston, 2003.

[Hru] E. Hrushovski, *The Mordell-Lang conjecture for function fields*, J. Amer. Math. Soc. **9**(3) (1996), 667-690.

[H-P] E. Hrushovski and A. Pillay, *Effective Bounds for the number of transcendental points on subvarieties of semi-abelian varieties*, Amer. J. Math. **122**(3) (2000), 439-450.

[Hun] T.W. Hungerford, *Algebra*, GTM 73, Springer-Verlag, New York, 1974.

[Kap_1] I. Kaplansky, *An Introduction to Differential Algebra*, 2^{nd}-edition, Hermann, Paris, 1976.

[Koc] H. Koch, *Algebraic Number Theory*, Springer-Verlag, Berlin, 1977.

[Kol_2] E.R. Kolchin, *Differential Algebra and Algebraic Groups*, Academic Press, New York, 1973.

[Lang] S. Lang, *Algebra*, Revised Third Edition, GTM 211, Springer-Verlag, New York, 2002.

[Mag] A.R. Magid, "Lectures on Differential Galois Theory," *University Lecture Series* **7**, American Mathematical Society, Providence, RI, 1994.

[N-Z-M] I. Niven, H.S. Zuckerman and H.L. Montgomery, *An Introduction to the Theory of Numbers*, Fifth Edition, John Wiley & Sons, New York, 1991.

[vdP-S] M. van der Put and M.F. Singer, *Galois Theory of Linear Differential Equations*, Springer-Verlag, Berlin, 2003.

[Scan] T. Scanlon, Model Theory and Differential Algebra, in *Differential Algebra and Related Topics*, Eds., L. Guo, P.J. Cassidy, W.F. Keigher & W.Y. Sit, World Scientific, Singapore, 2002.

[Serre] J.P. Serre, *A Course in Arithmetic*, GTM 7, Springer-Verlag, NY, 1973.