REMARKS ABOUT EUCLIDEAN DOMAINS

KEITH CONRAD

1. Introduction

The following definition of a Euclidean (not Euclidian!) domain is very common in textbooks. We write $\mathbf{N}$ for $\{0, 1, 2, \dots\}$.

Definition 1.1. An integral domain R is called *Euclidean* if there is a function $d\colon R-\{0\} \to \mathbf{N}$ with the following two properties:

(1) $d(a) \leq d(ab)$ for all nonzero a and b in R,
(2) for all a and b in R with $b \neq 0$ we can find q and r in R such that

$$a = bq + r, \quad r = 0 \text{ or } d(r) < d(b).$$

We will call (1) the *d-inequality*. Sometimes it is expressed in a different way: for nonzero a and b, if $a|b$ then $d(a) \leq d(b)$. This is equivalent to the d-inequality, taking into account of course the different roles of a and b in the two descriptions.

Examples of Euclidean domains are $\mathbf{Z}$ (with $d(n) = |n|$), $F[T]$ for any field F (with $d(f) = \deg f$; this example is *the* reason that one doesn't assume $d(0)$ is defined), and $\mathbf{Z}[i]$ ($d(\alpha) = \mathrm{N}(\alpha)$). A possibly silly kind of example is any field F with $d(a) = 1$ for all $a \neq 0$ (here all remainders are 0).

Definition 1.1 is in [1, §2.1], [3, §37], [4, §7.2], [5, §6.5], [7, §3.7], [8, Chap. III §3], [10, §12.4], [12, §3.6], and [13, §18]. However, in [2, §8.1] we find a different definition, where the d-inequality is missing:

Definition 1.2. An integral domain R is called *Euclidean* if there is a function $d\colon R-\{0\} \to \mathbf{N}$ such that for all a and b in R with $b \neq 0$ we can find q and r in R such that

$$a = bq + r, \quad r = 0 \text{ or } d(r) < d(b). \tag{1.1}$$

Any function $d\colon R - \{0\} \to \mathbf{N}$ which satisfies (1.1) will be called a *Euclidean function* on R. Thus a Euclidean domain in Definition 1.2 is an integral domain which admits a Euclidean function, while a Euclidean domain in Definition 1.1 is an integral domain which admits a Euclidean function satisfying the d-inequality.

Does Definition 1.2 describe a larger class of rings than Definition 1.1? No. We will show in Section 2 that any Euclidean domain (R, d) in the sense of Definition 1.2 can be equipped with a different Euclidean function $\widetilde{d}$ such that $\widetilde{d}(a) \leq \widetilde{d}(ab)$ for all a and b in R, so $(R, \widetilde{d})$ is Euclidean in the sense of Definition 1.1.

The main reason that the d-inequality is not included in the definition of a Euclidean domain in [2] is that it is irrelevant to prove the two main theorems about Euclidean domains: that every Euclidean domain is a PID and that the Euclidean algorithm in a Euclidean domain terminates after finitely many steps and produces a greatest common divisor. The Euclidean algorithm provides a practical method of computing greatest common divisors when it is available.

Why is the d-inequality nearly always mentioned in the (textbook) literature if it's actually not needed? Well, it is not needed for the two specific results cited in the previous paragraph, but it is convenient to use the d-inequality if we want to prove factorization into irreducibles in a Euclidean domain without proving the result more generally in a PID. There is factorization into irreducibles in any PID, which subsumes the same result for Euclidean domains since Euclidean domains are PIDs, but the proof of the existence of irreducible factorizations in a PID is less concrete than a proof available in Euclidean domains. We will see why in Sections 3 and 4.

In Section 5 we discuss Euclidean domains among quadratic rings.

2. Refining the Euclidean function

Suppose (R, d) is a Euclidean domain in the sense of Definition 1.2. We will introduce a new Euclidean function $\widetilde{d}\colon R - \{0\} \to \mathbf{N}$, built out of d, which satisfies $\widetilde{d}(a) \leq \widetilde{d}(ab)$. Then $(R, \widetilde{d})$ is Euclidean in the sense of Definition 1.1, so the rings which admit Euclidean functions in either sense are the same.

Here's the definition (trick?): for nonzero a in R, set

$$\widetilde{d}(a) = \min_{b \neq 0} d(ab).$$

That is, $\widetilde{d}(a)$ is the smallest d-value on the nonzero multiples of a. So $\widetilde{d}(a) = d(ab_0)$ for some b_0 and $d(ab_0) \leq d(ab)$ for all nonzero b. For example,

$$\widetilde{d}(1) = \min_{b \neq 0} d(b)$$

is the smallest d-value on $R - \{0\}$. An additional property is

$$\widetilde{d}(a) \leq d(a) \tag{2.1}$$

for all nonzero a in R, since a is a multiple of a.

Theorem 2.1. *Let (R, d) be a Euclidean domain in the sense of Definition* 1.2*. Then $(R, \widetilde{d})$ is Euclidean in the sense of Definition* 1.1*.*

Proof. For any nonzero a and b in R we have

$$\widetilde{d}(a) \leq \widetilde{d}(ab).$$

Indeed, write $\widetilde{d}(ab) = d(abc)$ for some nonzero c in R. Then, since abc is a multiple of a,

$$\widetilde{d}(a) \leq d(abc) = \widetilde{d}(ab).$$

We now show R admits a division theorem with respect to $\widetilde{d}$. Pick a and b in R with $b \neq 0$. Set $\widetilde{d}(b) = d(bc)$ for some nonzero $c \in R$. By the division theorem for (R, d) on the pair a and bc (not a and b) there are q_0 and r_0 in R such that

$$a = (bc)q_0 + r_0, \quad r_0 = 0 \text{ or } d(r_0) < d(bc).$$

Set $q = cq_0$ and $r = r_0$. Since $d(bc) = \widetilde{d}(b)$ and $\widetilde{d}(r) \leq d(r)$ (by (2.1)), the inequality $d(r) < d(bc)$ implies $\widetilde{d}(r) < \widetilde{d}(b)$. Thus

$$a = bq + r, \quad r = 0 \text{ or } \widetilde{d}(r) < \widetilde{d}(b).$$

Hence $(R, \widetilde{d})$ is a Euclidean domain in the sense of Definition 1.2. □

We end this section with a brief discussion of two other possible refinements one might want in a Euclidean function (but which we will not need later): uniqueness of the quotient and remainder it produces and multiplicativity.

In $\mathbf{Z}$ we write $a = bq + r$ with $0 \leq r < |b|$ and q and r are uniquely determined by a and b. There is also uniqueness of the quotient and remainder when we do division in $F[T]$ (relative to the degree function) and in any field (the remainder is always 0). Are there any other Euclidean domains where the quotient and remainder are unique? Division in $\mathbf{Z}[i]$ does not have a unique quotient and remainder relative to the norm on $\mathbf{Z}[i]$. For instance, dividing $1 + 8i$ by $2 - 4i$ gives

$$1 + 8i = (2 - 4i)(-1 + i) - 1 + 2i \text{ and } 1 + 8i = (2 - 4i)(-2 + i) + 1 - 2i,$$

where both remainders have norm 5, which is less than $\mathrm{N}(2 - 4i) = 20$.

Theorem 2.2. *If R is a Euclidean domain where the quotient and remainder are unique then R is a field or $R = F[T]$ for a field F.*

Proof. See [9] or [11]. □

This might be a surprise: $\mathbf{Z}$ isn't in the theorem! Aren't the quotient and remainder unique there? Yes if we take the remainder r so that $0 \leq r < |b|$, but not if we try to fit it into the setting of Euclidean domains using $|r| < |b|$. For instance,

$$51 = 6 \cdot 8 + 3 \text{ and } 51 = 6 \cdot 9 - 3.$$

The point is that using the Euclidan function on the remainder too, in the case of $\mathbf{Z}$, permits negative remainders so in fact $\mathbf{Z}$ does not have a unique quotient and remainder when we measure the remainder by its absolute value.

The Euclidean function in many basic examples satisfies a stronger property than $d(a) \leq d(ab)$, namely $d(ab) = d(a)d(b)$ with $d(a) \geq 1$ when $a \neq 0$. For instance, this holds in $\mathbf{Z}$ ($d(n) = |n|$) and $\mathbf{Z}[i]$ ($d(\alpha) = \mathrm{N}(\alpha)$). The degree on $F[T]$ is not multiplicative, but $d(f) = 2^{\deg f}$ is multiplicative. There is also a multiplicative Euclidean function on any field F: $d(a) = 1$ for all $a \neq 0$. Is there any Euclidean domain where the Euclidean function d can be chosen to satisfy the d-inequality but does not satisfy the stronger multiplicative property? We leave this as an open question.

3. Features of the d-inequality

Let R be a Euclidean domain. By Theorem 2.1 we may assume our Euclidean function d satisfies the d-inequality: $d(a) \leq d(ab)$ for all nonzero a and b in R. Using this inequality we will prove a few properties of d:

Theorem 3.1. *Let (R, d) be a Euclidean domain where d satisfies the d-inequality. Then*

1. *$d(a) \geq d(1)$ for all nonzero $a \in R$,*
2. *if $b \in R^\times$ then $d(ab) = d(a)$ for all nonzero a,*
3. *if $b \notin R^\times$ then $d(ab) > d(a)$ for all nonzero a.*

In particular, for nonzero a and b, $d(ab) = d(a)$ if and only if $b \in R^\times$.

Proof. (1): By the d-inequality, $d(1) \leq d(1 \cdot a) = d(a)$.

(2): By the d-inequality, $d(a) \leq d(ab)$. To get the reverse inequality, let c be the inverse of b, so the d-inequality implies

$$d(ab) \leq d((ab)c) = d(a).$$

(3): We want to show the inequality $d(a) \leq d(ab)$ is strict when b is not a unit. The proof is by contradiction. Assume $d(a) = d(ab)$. Now use division of a by ab:

$$a = (ab)q + r, \quad r = 0 \text{ or } d(r) < d(ab).$$

We rewrite this as

$$a(1 - bq) = r, \quad r = 0 \text{ or } d(r) < d(a).$$

Since b is not a unit, $1 - bq$ is nonzero, so $a(1 - bq)$ is nonzero. Thus $r \neq 0$, so the inequality $d(r) < d(a)$ becomes

$$d(a(1 - bq)) < d(a).$$

But this contradicts the d-inequality, which says $d(a) \leq d(a(1 - bq))$. Thus it is impossible for $d(a)$ to equal $d(ab)$ when b is not a unit. □

Note Theorem 3.1 is not saying two elements having the same d-value are unit multiples, but rather that a multiple has the same d-value if and only if it is a unit multiple. For example, in $\mathbf{Z}[i]$ we have $\mathrm{N}(1 + 2i) = \mathrm{N}(1 - 2i)$ but $1 + 2i$ and $1 - 2i$ are not unit multiples. Remember this!

Corollary 3.2. *Let (R, d) be a Euclidean domain where d satisfies the d-inequality. We have $d(a) = d(1)$ if and only if $a \in R^\times$. That is, the elements of least d-value in R are precisely the units.*

Proof. Take $a = 1$ in parts 2 and 3 of Theorem 3.1. □

We can see Corollary 3.2 working in $\mathbf{Z}$ and $F[T]$: the integers satisfying $|n| = |1|$ are ± 1, which are the units of $\mathbf{Z}$. The polynomials f in $F[T]$ satisfying $\deg f = \deg 1 = 0$ are the nonzero constants, which are the units of $F[T]$.

Corollary 3.3. *Let (R, d) be a Euclidean domain where d satisfies the d-inequality. If a and b are nonunits, then $d(a)$ and $d(b)$ are both less than $d(ab)$.*

Proof. This is immediate from part (3) of Theorem 3.1, where we switch the roles of a and b to get the inequality on both $d(a)$ and $d(b)$. □

4. Irreducible factorization

To see the simplicity introduced by a Euclidean function, we will prove irreducible factorization in both Euclidean domains and in PIDs.

Definition 4.1. Let R be an integral domain. A nonzero element a of R is called *irreducible* if it is not a unit and in any factorization $a = bc$, one of the factors b or c is a unit. A nonzero nonunit which is not irreducible is called *reducible.*

There are three types of nonzero elements in an integral domain: units (the invertible elements, whose factors are always units too), irreducibles (nonunits whose factorizations into two parts always involve one unit factor), and reducibles (nonunits which admit some factorization into a product of two nonunits). Notice that in a field there are no reducible or irreducible elements: everything is zero or a unit. So if we want to prove a theorem about irreducible factorization, we avoid fields.

Theorem 4.2. *In any Euclidean domain which is not a field, every nonzero nonunit is a product of irreducibles.*

Proof. Let (R, d) be a Euclidean domain which is not a field. By Theorem 2.1 we may assume $d(a) \leq d(ab)$ for all nonzero a and b in R. Therefore Corollary 3.3 applies.

We will prove the existence of irreducible factorizations by induction on the d-value. From Corollary 3.2, the units of R have the smallest d-value. Any $a \in R$ with second smallest d-value must be irreducible: if we write $a = bc$ and b and c are both nonunits, then $d(b)$ and $d(c)$ are both less than $d(a)$ by Corollary 3.3. Therefore b and c are units, so a is a unit. This is a contradiction.

Assume now that $a \in R$ is a nonunit and all nonunits with smaller d-value admit an irreducible factorization. To prove a admits an irreducible factorization too, we may suppose a is not irreducible itself. Therefore there is some factorization $a = bc$ with b and c both nonunits. Then $d(b) < d(a)$ and $d(c) < d(a)$ by Corollary 3.3, so b and c both have irreducible factorizations by induction. Thus their product a has an irreducible factorization. □

The conclusion of Theorem 4.2 is true for PIDs, but the proof will require more abstract methods than induction.

Theorem 4.3. *In a* PID *which is not a field, any nonzero nonunit is a product of irreducibles.*

The proof of Theorem 4.3 relies on the following lemma, which has a recursive flavor.

Lemma 4.4. *If R is an integral domain and $a \in R$ is a nonzero nonunit which does not admit a factorization into irreducibles then there is a strict inclusion of principal ideals $(a) \subset (b)$ where b is some other nonzero nonunit which does not admit a factorization into irreducibles.*

Proof. By hypothesis a is not irreducible, so (since it is neither 0 nor a unit either) there is some factorization $a = bc$ where b and c are nonunits (and obviously are not 0 either). If both b and c admitted irreducible factorizations then so does a, so at least one of b or c has no irreducible factorization. Without loss of generality it is b which has no irreducible factorization. Since c is not a unit, the inclusion $(a) \subset (b)$ is strict. □

Now we can prove Theorem 4.3.

Proof. Suppose there is an element a in the PID which is not 0 or a unit and has no irreducible factorization. Then by Lemma 4.4 there is a strict inclusion

$$(a) \subset (a_1)$$

where a_1 has no irreducible factorization. Then using a_1 in the role of a (and Lemma 4.4 again) there is a strict inclusion

$$(a_1) \subset (a_2)$$

where a_2 has no irreducible factorization. This argument (repeatedly applying Lemma 4.4 to the generator of the next larger principal ideal) leads to an infinite increasing chain of principal ideals

$$(a) \subset (a_1) \subset (a_2) \subset (a_3) \subset \cdots \tag{4.1}$$

where all inclusions are strict. This turns out to be impossible in a PID.

Indeed, suppose a PID contains an infinite strictly increasing chain of ideals:

$$I_0 \subset I_1 \subset I_2 \subset I_3 \subset \cdots$$

and set

$$I = \bigcup_{n \geq 0} I_n.$$

This union I is an ideal. The reason is that the I_n's are strictly increasing, so any *finite* set of elements from I lies in a common I_n. (This is the key idea.) So I is closed under addition and arbitrary multiplications from the ring since each I_n has these properties. (Make sure you understand that step.) Because we are in a PID, I is principal: $I = (r)$ for some r in the ring. But because I is the union of the I_n's, r is in some I_N. Then $(r) \subset I_N$ since I_N is an ideal, so

$$I = (r) \subset I_N \subset I,$$

which means

$$I_N = I.$$

But this is impossible because the inclusion $I_{N+1} \subset I$ becomes $I_{N+1} \subset I_N$ and we were assuming I_N was a proper subset of I_{N+1}. Because of this contradiction, nonzero nonunits in a PID without an irreducible factorization do not exist. □

The proof that a PID does not contain an infinite strictly increasing chain of ideals holds for a broader class of rings than PIDs.

Theorem 4.5. *A commutative ring in which every ideal is finitely generated does not contain an infinite strictly increasing chain of ideals.*

Proof. The second half of the proof of Theorem 4.3 works in this more general context. All we have to do is show the logic works when I is finitely generated rather than principal. The point is that if $I = (x_1, \dots, x_m)$ then the finitely many x_i's all lie in some common I_N (because the I_n's are an increasing chain). And now the contradiction is obtained just as before: $I_N = I$ but then $I_{N+1} \subset I_N$, contradiction. □

Corollary 4.6. *In an integral domain where every ideal is finitely generated, every nonzero nonunit has an irreducible factorization.*

Proof. If there were an element a which is not 0 or a unit and which did not admit an irreducible factorization then, as in the proof of Theorem 4.3, we could produce an infinite strictly increasing chain of (principal) ideals. But there are no infinite strictly increasing chains of ideals in the ring, by Theorem 4.5. □

Definition 4.7. A commutative ring where every ideal is finitely generated is called a *Noetherian* ring.

These rings are named after Emmy Noether, who was one of the pioneers of abstract algebra in the first half of the 20th century. Their importance, as a class of rings, stems from the stability of the Noetherian property under many basic constructions. If R is a Noetherian ring, so is any quotient ring R/I (which may not be an integral domain even if R is), any polynomial ring $R[X]$ (and thus $R[X_1, \dots, X_n]$ by induction on n, viewing this as $R[X_1, \dots, X_{n-1}][X_n]$), and any formal power series ring $R[[X]]$ (and thus $R[[X_1, \dots, X_n]]$). The PID property behaves quite badly, *e.g.*, if R is a PID other than a field then $R[X]$ is not a PID. For instance, $R[X, Y] = R[Y][X]$ is never a PID for any integral domain R (other than the zero ring). But if R is Noetherian then $R[X, Y]$ is Noetherian. Briefly, the property "ideals are finitely generated" of Noetherian rings is much more flexible than the property "ideals are singly generated" of PIDs.

Using this terminology, Corollary 4.6 says in any *Noetherian integral domain* any element other than 0 or a unit has an irreducible factorization. It is worth comparing the proof of this general result (Corollary 4.6) to the special proof we gave in the case of Euclidean domains, where the proof of irreducible factorizations is tied up with features of the Euclidean function on the ring.

In the context of unique factorization domains, it is the uniqueness of the factorization which lies deeper than the existence. We are not discussing uniqueness here, which most definitely does *not* hold in most Noetherian integral domains. That is, the existence of irreducible factorizations (for all nonzero nonunits) is not a very strong constraint, to the extent that most integral domains you meet in day-to-day practice in mathematics are Noetherian so their elements automatically have some factorization into irreducible elements. But there usually is not going to be a unique factorization into irreducible elements.

5. Euclidean and non-Euclidean quadratic rings

The main importance of Euclidean domains in an algebra course is a ready source of examples of PIDs. Three points are worth noting:

- Aside from computational issues (as in the Euclidean algorithm) it is more useful to know whether or not a ring is a PID than whether or not it is Euclidean.
- There are PIDs which are not Euclidean.
- Even if an integral domain is Euclidean, there are methods (for certain kinds of rings) which let one show the ring is a PID while bypassing the question of whether or not it is Euclidean. For example, the ring $\mathbf{Z}[\sqrt{14}]$ was known to be a PID since the 19th century (definitely since the 1870s, although from a different point of view this was even known to Gauss at the beginning of the 1800s). It was proved to be Euclidean only in the 21st century [6].

The simplest setting where one can find PIDs which are not Euclidean are found among the quadratic rings.

Definition 5.1. A *quadratic ring* is a ring of the form $\mathbf{Z}[\gamma]$ where γ is the root of a monic irreducible quadratic polynomial $T^2 + aT + b \in \mathbf{Z}[T]$. We call $\mathbf{Z}[\gamma]$ *real* if γ is real and *imaginary* otherwise.

For instance, the Gaussian integers $\mathbf{Z}[i]$ are an imaginary quadratic ring (associated to the polynomial $T^2 + 1$). The ring $\mathbf{Z}[(1+\sqrt{5})/2]$ is real quadratic, with $(1+\sqrt{5})/2$ a root of $T^2 - T - 1$. By the quadratic formula, γ has at worst a 2 in the denominator. Note the real quadratic rings are subrings of $\mathbf{R}$ (with $a^2 - 4b > 0$) and the imaginary quadratic rings are not ($a^2 - 4b < 0$).

Since $\gamma^2 = -a\gamma - b \in \mathbf{Z} + \mathbf{Z}\gamma$, by induction every power of γ is in $\mathbf{Z} + \mathbf{Z}\gamma$, so

$$\mathbf{Z}[\gamma] = \mathbf{Z} + \mathbf{Z}\gamma.$$

We can't necessarily "complete" the square and write a quadratic ring in the form $\mathbf{Z}[\sqrt{m}]$ for some m. The point is the root of T^2+aT+b can have a denominator of 2, which can't be absorbed by the coefficients when working over $\mathbf{Z}$ (rather than, say, over $\mathbf{Q}$). For instance, $\mathbf{Z}[(1+\sqrt{5})/2] \neq \mathbf{Z}[\sqrt{m}]$ for any m.

When γ is one root of $T^2 + aT + b$, the other root is $\overline{\gamma} = -a - \gamma$, which is called the *conjugate* of γ. More generally, for $\alpha = x + y\gamma$ in $\mathbf{Z}[\gamma]$, its *conjugate* is taken to be

$$\overline{\alpha} := x + y\overline{\gamma} = x - ay - y\gamma.$$

In the special case that $a = 0$ and we write $b = -m$, so γ is a root of $T^2 - m$ (a square root of m), we have $\overline{\gamma} = -\gamma$ and $\overline{x + y\gamma} = x - y\gamma$. The *norm* of α is defined to be

$$\mathrm{N}(\alpha) = \alpha\overline{\alpha} = x^2 - axy + by^2.$$

This is an integer, and it is nonzero unless $\alpha = 0$. Notice its coefficients do not coincide exactly with those of the polynomial $T^2 + aT + b$; the a occurs with the opposite sign (which only matters when $a \neq 0$). When $\alpha = c$ is in $\mathbf{Z}$ then $\mathrm{N}(c) = c^2$. In particular, $\mathrm{N}(\pm 1) = 1$.

A direct calculation shows the norm is multiplicative:

$$\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta).$$

Example 5.2. If $\gamma = \sqrt{2}$ then $\mathrm{N}(x + y\sqrt{2}) = x^2 - 2y^2$, which takes both positive and negative values, *e.g.*, $\mathrm{N}(3 + 5\sqrt{2}) = -41$.

Example 5.3. If $\gamma = \sqrt{m}$ then $\mathrm{N}(x + y\gamma) = x^2 - my^2$. This has both positive and negative values if $m > 0$ and only nonnegative values if $m < 0$.

Example 5.4. If $\gamma = (1 + \sqrt{5})/2$, a root of $T^2 - T - 1$, then $\mathrm{N}(x + y\gamma) = x^2 + xy - y^2$.

Several "small" quadratic rings are Euclidean with the absolute value of the norm as a Euclidean function. For instance, $\mathbf{Z}[i]$, $\mathbf{Z}[\sqrt{2}]$, and $\mathbf{Z}[\sqrt{-2}]$ are all Euclidean using $d(\alpha) = |\,\mathrm{N}(\alpha)|$. (For an imaginary quadratic ring like $\mathbf{Z}[i]$ we can drop the absolute value sign: the norm is already nonnegative.) This leads to two questions about a quadratic ring: is it Euclidean (with respect to some Euclidean function) and is it norm-Euclidean (*i.e.*, Euclidean using the particular choice $d(\alpha) = |\,\mathrm{N}(\alpha)|$)? To prove a quadratic ring is not Euclidean, it is not sufficient to show the absolute value of the norm can't work; maybe there is a different Euclidean function available. Indeed, $\mathbf{Z}[\sqrt{14}]$ is such an example. It was recently proved to be Euclidean but it has been known not to be norm-Euclidean since the early 20th century.

The rest of this handout is concerned with setting up the background to prove a particular imaginary quadratic ring which is a PID is not Euclidean.

Theorem 5.5. *In a quadratic ring* $\mathbf{Z}[\gamma]$*, the units are the elements with norm* ± 1.

Proof. If $\alpha\beta = 1$ in $\mathbf{Z}[\gamma]$ then taking norms of both sides shows $\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = \mathrm{N}(1) = 1$ in $\mathbf{Z}$, so $\mathrm{N}(\alpha) = \pm 1$. Conversely, if $\mathrm{N}(\alpha) = \pm 1$ then $\alpha\overline{\alpha} = \pm 1$, so α is invertible (with inverse $\pm\overline{\alpha}$). □

Example 5.6. The units of $\mathbf{Z}[\sqrt{2}]$ are built from integral solutions to $x^2 - 2y^2 = \pm 1$. For instance, one solution is $x = 1$ and $y = 1$, giving the unit $1 + \sqrt{2}$. Its powers are also units (units are closed under multiplication), so $\mathbf{Z}[\sqrt{2}]$ has infinitely many units.

Example 5.7. Units in $\mathbf{Z}[\sqrt{3}]$ come from integral solutions to $x^2 - 3y^2 = \pm 1$. However, there are no solutions to $x^2 - 3y^2 = -1$ since the equation has no solutions modulo 3: $x^2 \equiv -1 \bmod 3$ has no solution. Thus the units of $\mathbf{Z}[\sqrt{3}]$ only correspond to solutions to $x^2 - 3y^2 = 1$. One nontrivial solution (that is, other than ± 1) is $x = 2$ and $y = 1$, which yields the unit $2 + \sqrt{3}$. Its powers give infinitely many more units.

Example 5.8. The units of $\mathbf{Z}[\sqrt{-2}]$ come from integral solutions to $x^2 + 2y^2 = 1$. The right side is at least 2 once $y \neq 0$, so the only integral solutions are $x = \pm 1$ and $y = 0$, corresponding to the units ± 1. In contrast to the previous two examples, where there are infinitely many units, $\mathbf{Z}[\sqrt{-2}]$ has only two units.

The following theorem about Euclidean domains is the key to proving certain (imaginary) quadratic rings are not Euclidean. Notice the proof does not require the Euclidean function on the ring to satisfy the d-inequality.

Theorem 5.9. *Let (R, d) be a Euclidean domain. Let $a \in R$ be a nonunit with least d-value among all non-units. Then the quotient ring $R/(a)$ is represented by 0 and units.*

Proof. Pick any $x \in R$. By the division theorem we can write $x = aq + r$ where $r = 0$ or $d(r) < d(a)$. If $r \neq 0$, then the inequality $d(r) < d(a)$ forces r to be a unit. Since $x \equiv r \bmod a$, we conclude that $R/(a)$ is represented by 0 and by units. □

Example 5.10. When $R = \mathbf{Z}$ we can use $a = 2$. Then $\mathbf{Z}/(2)$ is represented by 0 and 1. When $R = \mathbf{Z}[i]$ we can use $a = 1 + i$. Then $\mathbf{Z}[i]/(1 + i)$ is represented by 0 and 1. These examples show some units could be congruent modulo a, but at least every element of the ring is congruent to 0 or some (perhaps more than one) unit.

We have shown that if R is a Euclidean domain, there are elements of R (namely nonunits with least d-value) modulo which everything is congruent to 0 or to a unit from R. A ring where no element has this property therefore can't be a Euclidean domain.

Remark 5.11. In a domain R, an element a for which the ring $R/(a)$ is represented by 0 and units in R is called a *universal side divisor* in the literature. This terminology seems stupid. Side divisor? What's that? Remember the property, but forget the label (and don't ever use it, because nobody will know what you're talking about).

Example 5.12. The quadratic ring $R = \mathbf{Z}[(1 + \sqrt{-19})/2]$ is a PID which is not Euclidean. This is proved in [2, pp. 277,282]. We will include here the proof that R is not Euclidean by showing R contains no element a for which $R/(a)$ is represented by 0 and units.

First we compute the norm of a typical element $\alpha = x + y(1 + \sqrt{-19})/2$:

$$\mathrm{N}(\alpha) = x^2 + xy + 5y^2 = \left(x + \frac{y}{2}\right)^2 + \frac{19y^2}{4}. \tag{5.1}$$

This norm always takes values ≥ 0 (this is clearer from the second expression for it than the first) and once $y \neq 0$ we have $\mathrm{N}(\alpha) \geq 19y^2/4 \geq 19/4 > 4$. In particular, the units are solutions to $\mathrm{N}(\alpha) = 1$, which are ± 1:

$$R^\times = \{\pm 1\}.$$

The first few norm values are 0, 1, 4, 5, 7, and 9. In particular, there is no element of R with norm 2 or 3. This and the fact that $R^\times \cup \{0\}$ has size 3 are the key facts.

If R were Euclidean then there would be a nonunit a in R such that $R/(a)$ is represented by 0 and units, so by 0, 1, and -1. Perhaps $1 \equiv -1 \bmod a$, but we definitely have $\pm 1 \not\equiv 0 \bmod a$ (since a is not a unit, (a) is a proper ideal so $R/(a)$ is not the zero ring). Thus $R/(a)$ has size 2 (if $1 \equiv -1 \bmod a$) or 3 (if $1 \not\equiv -1 \bmod a$). We show this can't happen.

If $R/(a)$ has size 2 then $2 \equiv 0 \bmod a$ (think about $R/(a)$ as an additive group of size 2), so $a|2$ in R. Therefore $\mathrm{N}(a)|4$ in $\mathbf{Z}$. There are no elements of norm 2, so the only nonunits with norm dividing 4 are elements with norm 4. A check using (5.1) shows the only such numbers are ± 2. However, $R/(2) = R/(-2)$ does not have size 2. For instance, 0, 1, and $(1 + \sqrt{-19})/2$ are incongruent modulo ± 2: the difference of any two of these, divided by two, is not of the form $x + y(1 + \sqrt{-19})/2$ for x and y in $\mathbf{Z}$.

Similarly, if $R/(a)$ has size 3 then $a|3$ in R, so $\mathrm{N}(a)|9$ in $\mathbf{Z}$. Since there is no element with norm 3, a must have norm 9 (it doesn't have norm 1 since it is not a unit). The only elements of R with norm 9 are ± 3, so $a = \pm 3$. The ring $R/(3) = R/(-3)$ does not have size 3: 0, 1, 2, and $(1+\sqrt{-19})/2$ are incongruent modulo ± 3 since their differences divided by 3 are not in $\mathbf{Z}[(1+\sqrt{-19})/2]$.

Since $\#(R^\times \cup \{0\}) = 3$ and R has no element a such that $R/(a)$ has size 2 or 3, R can't be a Euclidean domain.

References

[1] S. Alaca and K. S. Williams, "Introductory Algebraic Number Theory," Cambridge Univ. Press, 2003.
[2] D. Dummit and R. Foote, "Abstract Algebra," 3rd ed., Wiley, 2004.
[3] J. Durbin, "Modern Algebra: An Introduction," 5th ed., J. Wiley, 2004.
[4] J. B. Fraleigh, "A First Course in Abstract Algebra," 6th ed., Addison-Wesley, 1999.
[5] F. Goodman, "Algebra: Abstract and Concrete (Stressing Symmetry)," 2nd ed., Prentice-Hall, 2005.
[6] M. Harper, $\mathbf{Z}[\sqrt{14}]$ is Euclidean, *Canad. J. Math.* **56** (2004), 55–70.
[7] I. Herstein, "Topics in Algebra," 2nd ed., Wiley, 1975.
[8] T. Hungerford, "Algebra," Springer-Verlag, 1980.
[9] M. A. Jodeit, Uniqueness in the division algorithm, *Amer. Math. Monthly* **74** (1967), 835–836.
[10] C. Lanski, "Concepts in Abstract Algebra," Brooks/Cole, 2005.
[11] T. S. Rhai, A characterization of polynomial domains over a field, *Amer. Math. Monthly* **69** (1962), 984–986.
[12] J. Rotman, "Advanced Modern Algebra," Prentice-Hall, 2002.
[13] B. L. van der Waerden, "Modern Algebra," Ungar, New York, 1953.