Notes

101.26 Primes of the form $x^2 + z^2$

Fermat's Theorem that a prime congruent to 1 mod 4 can be expressed as the sum of two squares has been extremely important in motivating the initial development of class field theory.

A relatively short solution was published by D. R. Heath-Brown in 1984 in *Invariant*, an occasional publication of the Oxford Invariants Society. Since then, several variations of this have appeared and Don Zagier's proof, [1], has brought the idea to the attention of many more people. However, the solution which follows is more elementary. It was devised to be an example of proof by contradiction for undergraduates. It also illustrates the useful stratagem of counting a quantity in two different ways.

Let $p \equiv 1 \bmod 4$ be prime and suppose there are no integer solutions of $x^2 + z^2 = p$.

Consider integer solutions of $4xy + z^2 = p$ with x and y positive. Note that x and y cannot be equal and z can be neither zero nor $y - x$. Given any solution, one can interchange x and y and/or change the sign of z. The number of solutions is therefore a multiple of 4. Precisely half of these, an even number, have $x - y + z > 0$.

Now consider integer solutions of $4xy + z^2 = p$ with x, y and $x - y + z$ positive. Then (a, b, c) is a solution if, and only if, $(a - b + c, b, 2b - c)$ is a solution. For two such solutions to be equal, $b = c$ and $b(4a + b) = p$. Therefore $b = 1$ and $4a + 1 = p$. The only possibility is $\left(\frac{1}{4}(p - 1), 1, 1\right)$ and so the number of solutions is odd.

This contradiction proves that there *is* an integer solution of $x^2 + z^2 = p$.

References

1. Don Zagier, A one-sentence proof that every prime $p \equiv 1 \bmod 4$ is a sum of two squares, *Amer. Math. Monthly* **97** (1990), no. 2, p. 144.

10.1017/mag.2017.128

STAN DOLAN

126A Harpenden Road, St Albans AL3 6BZ

101.27 On the convention that $\deg(0) = -\infty$

The ring of polynomials in one variable x over any integral domain D (such as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, etc.) is denoted by $D[x]$. Every *non-zero* polynomial in $D[x]$ can be written uniquely in the form

$$f = a_0 + a_1 x + \ldots + a_n x^n, n \geqslant 0, a_0, a_1, \ldots, a_n \in D[x], a_n \neq 0. \qquad (1)$$

For such an f, n is called the *degree* of f and is written as $\deg(f)$. Since the zero polynomial 0 cannot be written in the form (1), $\deg(0)$ may be considered undefined. This seems to be the convention made in most undergraduate textbooks. These include [1, p. 346], [2, p. 285], [3, p. 154], [4, p. 236], [5, p. 97] and [6, p. 157]. In other books, e.g. [7, p. 234], the

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press

degree of a polynomial f is defined if f is not the zero polynomial, but nothing is mentioned about the degree of the zero polynomial, seemingly leaving it to the readers to make their own decisions. The author of [8, p. 350] states that *the degree of the zero polynomial is indeterminate*, and the author of [9, p. 166] states that *the zero polynomial has all degrees*. This may seem odd, but it may be the only way out of the situation described in (c) below. Other books define $\deg(0)$ to be 0. This is explicit in [10, p. 265] and follows from a careful reading of [11, p. 54]; it is also implicit in [12, p. 46], where the author identifies $\{f \in D[x] : \deg(f) = 0\}$ with D. In [13, p. 329], the authors do not prescribe any degree to the zero polynomial, and they add that, sometimes for the sake of convenience, the zero element is considered as a polynomial of infinite degree. Most graduate textbooks seem to adopt the convention that

$$\deg(0) = -\infty. \tag{2}$$

These include [14, A.IV.3], [15, p. 149], [16, p. 158], [17, p. 100] and [18, p. 21]. Some of these books, e.g. [15, p. 149] and [18, p. 63], even give reasons for this convention. These reasons are included in the discussion below. However, none of these books observes any of the three drawbacks of (2) discussed in (a), (b) and (c) below.

This Note is concerned with the convention (2). It gives a list of five reasons why such a convention is useful, and a list of three contexts in which it has drawbacks. We start with the first list.

(i) One way of circumventing the difficulty encountered in defining $\deg(0)$ at the beginning of this Note is to modify (1) above by dropping the requirement $a_n \neq 0$, i.e., to write f in the form

$$f = a_0 + a_1x + \ldots + a_nx^n,\ n \geqslant 0,\ a_0, \ldots, a_n \in D,$$

and to then define

$$\deg(f) = \sup\{k \geqslant 0 : a_k \neq 0\}.$$

This coincides with the previous definition if $f \neq 0$. However, when $f = 0$, the set $\{k \geqslant 0 : a_k \neq 0\}$ is the empty set $\varnothing$. Since the convention that

$$\sup(\varnothing) = -\infty \tag{3}$$

is reasonable and universally accepted, we are now forced to make the convention (2). We shall elaborate on the reasonableness of (3) later in this note.

(ii) The formula $\deg(fg) = \deg(f) + \deg(g)$ holds for all non-zero f and g. If we wish it to hold for *all* f and g, i.e. when one (or both) of the polynomials f and g is (or are) zero, then we need to have

$$\deg(0) = \deg(0) + n \text{ for all } n \geqslant 0 \text{ and also for } n = \deg(0).$$

The only way for this to hold is to have

$$\deg(0) = \pm\infty,$$

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press

so that the convention (2) is appropriate for this context. Here and throughout, the usual formal conventions about arithmetic involving ∞ and $-\infty$ are adopted.

(iii) The inequality

$$\deg(f + g) \leqslant \max\{\deg(f), \deg(g)\} \tag{4}$$

holds for all non-zero f and g. If we wish it to hold for *all* f and g, then, applying it to f and $-f$, we need to have

$$\deg(0) \leqslant \max\{\deg(f), \deg(-f)\} = \deg(f) \text{ for every } f.$$

Thus we must have

$$\deg(0) \leqslant 0, \tag{5}$$

and convention (2) accords with (5). This requirement does also accommodate the inequality

$$\deg(f) < \deg(g) \Rightarrow \deg(f + g) = \deg(g), \tag{6}$$

which supplements (4), because otherwise $\deg(1) = 0 < \deg(0)$ gives the contradiction $\deg(1 + 0) = \deg(0)$; see [14, A.IV.3, Proposition 1 (i)].

(iv) The requirement (5) is also needed if we want the set

$$\{f \in D[x] : \deg(f) \leqslant n\}, n \geqslant 0$$

to be an additive subgroup (or even a D-submodule) of $D[x]$.

(v) When D is a field (such as $\mathbb{Q}$ or $\mathbb{R}$), the Euclidean division property holds in $D[x]$. Thus if f and g are polynomials such that $g \neq 0$, then there exist polynomials q and r such that

$(*)$ $f = gq + r$,

$(**)$ $\deg(r) < \deg(g)$ or $r = 0$.

If we wish $(**)$ to have the two conditions $[\deg(r) < \deg(g)]$ and $[r = 0]$ merged so that $(**)$ takes the more elegant form

$(***)$ $\deg(r) < \deg(g)$,

then we would need to have

$$\deg(0) < 0.$$

So, with convention (2), the proof of the Euclidean division property would go by induction as usual, but instead of using the well-ordering property of $\mathbb{N}$ or $\mathbb{Z}_{\geqslant 0}$, one would now use the well ordering property of $\mathbb{Z}_{\geqslant 0} \cup \{-\infty\}$. However, the most convenient way for proving all the facts above is probably to treat the case of zero polynomials separately.

Summarising, (i) to (v) above are simultaneously satisfied when convention (2) holds. The majority of university and college students taking

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press

a first course in abstract algebra find (2) uncomfortable and even irritating. But some of the better students may enjoy thinking of the considerations above as a *proof* and of (2) as a *theorem*, and may look for other examples that support (2). Some may be curious whether there are statements that hold for non-zero polynomials, but would not hold for the zero polynomial if we accept (2). This author would like to hear from anyone regarding such examples, and he steps forward with three examples that are not in much harmony with (2).

(a) If we are to accept the statement that

the number of zeros (or roots) of a polynomial
is less than or equal to its degree, (7)

then we would have to set $\deg(0) = \infty$, and not $-\infty$.

(b) If f' denotes the derivative of f, then the formula

$$\deg(f') = \deg(f) - 1 \tag{8}$$

holds for non-constant polynomials f. If f is a non-zero constant, then (8) gives $\deg(0) = -1$, but this contradicts (8) applied to the zero polynomial.

(c) Some books, such as [13, p. 329], state that we can write any polynomial f in several variables as a sum of homogeneous polynomials of different degrees, say

$$f = f_0 + f_1 + \ldots + f_k,$$

where $k = \deg f$, and

$$f_i \text{ is homogeneous of degree } i. \tag{9}$$

For this to hold, one should declare that the zero polynomial has all degrees, as done in [9, p. 166]. A better remedy is to replace (9) by the accurate statement

$$f_i = 0 \text{ or } f_i \text{ is homogeneous of degree } i. \tag{10}$$

Beginners find both conventions (3) and (2) hard to swallow. But one can see that (3) follows from the definition of the supremum and from the fact that every real number r (including negative ones with large absolute value) is an upper bound of $\varnothing$. This in turn follows from the implication

$$t \in \varnothing \Rightarrow t \leqslant r \tag{11}$$

which is valid for all real numbers t and r. But here again, many students do not find (11) comfortable either, and need to be reminded that it follows from the *fact* or *axiom* that

if p is a false statement, then the implication $p \Rightarrow q$
is valid for all statements q.

Thus

if you start with a false statement,
then you can prove anything you like, (12)

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press

a claim that, in turn, may not be easier to swallow than any of the previous conventions (3), (2), and (11).

Finally, it would be interesting to run some kind of statistics to see which of the conventions (3), (2), (11) and (12) is most (or least) horrible! It would also be interesting to know how popular the convention that deg (0) be left undefined is.

Acknowledgement: I am grateful to the referee for his valuable suggestions.

References

1. D. M. Burton, *Abstract algebra*, Wm. C. Brown Publishers, Dubuque, Iowa (1988).
2. J. A. Gallian, *Contemporary abstract algebra* (5th edn.), Houghton Mifflin Company, Boston (2002).
3. I. N. Herstein, *Topics in algebra* (2nd edn.), Wiley Eastern Limited, New Delhi (1975).
4. I. M. Isaacs, *Algebra: a graduate course*, Brooks/Cole Publishing Company, Pacific Grove, California (1994).
5. N. H. McCoy, *Introduction to Modern Algebra* (4th edn.), (revised by Gerald J. Janusz), Allyn and Bacon Inc., Boston (1987).
6. J. Rotman, *A first course in abstract algebra*, Prentice Hall, Upper Saddle River (1996).
7. D. S. Dummit and R. M. Foote, *Abstract algebra*, John Wiley & Sons, New York (2004).
8. M. Artin, *Algebra*, Prentice Hall, Englewood Cliffs, New Jersey (1991).
9. P. M. Cohn, *Basic algebra: groups, rings, and fields*, Springer, London (2003).
10. J. B. Fraleigh, *A first course in abstract algebra* (3rd edn.), Addison-Wesley Publishing Company, Reading, Massachusetts (1982).
11. G. Birkhoff and S. MacLane, *A survey of modern algebra* (3rd edn.), The MacMillan Company, Reading, New York (1965).
12. B. L. van der Waerden, *Modern algebra*, Volume I, Frederick Ungar Publishing Co., New York (1966).
13. M. R. Dixon, L. A. Kurdachenko and I. Ya. Subbotin, *Algebra and number theory*, Wiley, New Jersey 2010.
14. N. Bourbaki, *Elements of mathematics, algebra II*, Springer-Verlag, New York (1990) Chapters 4-7.
15. P. M. Cohn, *Algebra*, Volume 1 (2nd edn.), John Wiley & Sons, New York (1989).
16. T. W. Hungerford, *Algebra*, Springer-Verlag, New York (1974).
17. S. Lang, *Algebra* (3rd edn.), Addison-Wesley Publishing Company, Reading, Massachusetts, New York, Upper Saddle River (1993).

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press

18. W. J. Wickless, *A first graduate course in abstract algebra*, Marcel Dekker, Inc., New York (2004).

10.1017/mag.2017.129

MOWAFFAQ HAJJA
Department of Basic Sciences and Mathematics, Faculty of Science, Philadelphia University, PO Box 1, 19392, Amman, Jordan
e-mail: *mowhajja@yahoo.com*

101.28 The versatile exponential inequality $e^x \geqslant 1 + x$

Introduction

In this Note, we survey and gather together a variety of applications of the inequality $e^x \geqslant 1 + x$. Many of them are scattered throughout the existing literature, but *en masse* they exemplify Bromwich's premise in [1] that "These simple inequalities are often sufficient to obtain useful properties of the exponential function.". The inequality

$$e^x \geqslant 1 + x \text{ for all real } x, \text{ with equality if, and only if, } x = 0 \qquad (1)$$

expresses the fact that the graph of $y = e^x$ lies above its tangent line at $x = 0$ (indeed, at any point $x = a$, since $e^{x-a} \geqslant 1 + x - a$ leads to $e^x \geqslant e^a + e^a(x - a)$). There are many different proofs of (1): we give two, probably less familiar ones, adapted from David Bradley's article [2] on the corresponding inequality for $\ln x$.

The first is the 'one-liner'

$$e^x - x - 1 = \begin{cases} \int_0^x \int_0^y e^t \, dt \, dy, & x \geqslant 0 \\ \int_x^0 \int_y^0 e^t \, dt \, dy, & x < 0 \end{cases}$$

where the integrands are manifestly non-negative.

For the second,

$$\begin{aligned} 0 \leqslant \sum_{n=1}^{N} 2^{n-1}\left(e^{x2^{-n}} - 1\right)^2 &= \sum_{n=1}^{N} \left(2^{n-1}e^{x2^{1-n}} - 2^n e^{x2^{-n}} + 2^{n-1}\right) \\ &= e^x - 2^N e^{x2^{-N}} + 2^N - 1 \text{ (by telescoping)} \\ &= e^x - 1 - \frac{x\left(e^{x2^{-N}} - 1\right)}{x2^{-N}} \end{aligned}$$

where the right-hand side tends to $e^x - 1 - x$ as $N \to \infty$ because e^x has derivative 1 at $x = 0$.

We will also use two immediate consequences of (1). First,

$$e^x \leqslant \frac{1}{1 - x} \text{ for } x < 1, \text{ with equality if, and only if, } x = 0. \qquad (2)$$

For (1) gives $e^{-x} \geqslant 1 - x$ which then gives (2) for the range of values quoted. We also have

$$e^x \geqslant ex \text{ for all real } x, \text{ with equality if, and only if, } x = 1. \qquad (3)$$

https://doi.org/10.1017/mag.2017.129 Published online by Cambridge University Press